Kreiner-Moller, E., Waage, J., Standl, M., Brix, S., Pers, T. H., Alves, A. C., Warrington, N. M., Tiesler, C. M. T., Fuertes, E., Franke, L., Hirschhorn, J. N., James, A., Simpson, A., Tung, J. Y., Koppelman, G. H., Postma, D. S., Pennell, C. E., Jarvelin, M-R., Custovic, A., ... Bønnelykke, K. (2017). Shared genetic variants suggest common pathways in allergy and autoimmune diseases. *Journal of Allergy and Clinical Immunology*. https://doi.org/10.1016/j.jaci.2016.10.055

Supplement - Shared genetic variants suggest common pathways in allergy and autoimmune diseases

Supplementary Methods

Discovery and meta-analysis

Discovery results from an allergic sensitization GWAS was included as one of the datasets comprising results from 5,809 with allergic sensitization and 9,875 controls with data from the following cohorts: AAGC, ALSPAC[1], B58C, COPSAC2000, LISA, MAAS, NFBC 1966, RAINE and PIAMA. Please see discovery paper for ethical statements, cohort profiles and numbers.[2] Sensitization status was assessed objectively by either elevated levels of allergen-specific IgE (sIgE) in blood or by a positive skin reaction after skin prick test (SPT) against common food and inhalant allergens. The SPT cut off level was 3 mm larger than the negative control for cases, and below 1 mm for controls. For sIgE, cases were defined as levels at or above 3.5 IU/mL and for controls below 0.35 IU/mL. Imputation was independently conducted for each study with reference to HapMap phase 2 or 3 CEU genotypes (Central European ancestry). Study level association analysis was performed using logistic regression models based on an expected additive allelic dosage model for SNPs, adjusting for ancestry-informative principal components as necessary. SNPs with MAF <1% and/or poor imputation quality (MACH r^2 <0.3 or IMPUTE proper info <0.4) were excluded. After genomic control at the level of the individual studies, the summary statistics were meta-analyzed using a fixed effects model and inverse variance as weights in METAL (2010-08-01). In total 2,400,129 SNPs were available in three or more cohorts.

GWAS results from the 23andMe study were included in the present study involving 10,509 individuals with self-reported cat-allergy, 9,815 with Dust-mite allergy, 16,133 with Pollen allergy (grasses, trees or weeds) and 26,311 without symptoms in a total of 46,646 individuals. Allergy was defined as those individuals who reported a positive allergy test, difficulty in swallowing or speaking, hives, itchy mouth, itchy eyes, itchy nose or asthma in response to a particular allergen.[3] The second part of the study sample on self-reported allergy (mothers from the Alspac cohort) was not included in the present study as these individuals are related to individuals in the GWA on sensitization.

Imputation was performed in the 23andMe study using the 1000 Genomes reference (August 2010 release) in batches. SNPs with an imputed $r^2 > 0.5$ averaged across all batches and $r^2 > 0.3$ in every batch were used. SNPs were remapped to B36. A generalized estimating equation (GEE) model was applied to assess the shared genetic effects on all three phenotypes taking into account the correlations between these phenotypes.

The meta-analysis was performed using a fixed effects model using inverse variance as weights in METAL (2010-08-01)[4] after a second genomic control for the meta-analysis of the dataset on self reported allergy and sensitization.

Enrichment of autoimmune disease-associated loci and allergy

Candidate loci were chosen from the GWA's catalog[5] accessed 25th of November 2013 using autoimmune and inflammatory traits of interest (**Supplementary Table E1**). These reported traits were collapsed into 16 overall autoimmune disease traits. All SNP-trait associations with P<5e-8 were used. We collapsed close SNPs into loci (+/-250kb)[6] and used for each locus the SNP with lowest reported P as index SNP (**Supplementary Figure 1**). For common loci (listed in table 2), all original publications were checked for effect allele, and any discrepancies with the GWA'S catalog was corrected. After the extraction of SNP-associations, the enrichment Odds Ratio was calculated as the number of observed extracted SNPs with P<0.05 out of total extracted SNPs as compared to total number of independent SNPs with P<0.05 within the GWA discovery results using a Fisher's exact test. For this we used Hapmap, CEU panel, to define independent loci. This was performed using PLINK (--indep-pairwise 200 5 0.5) with a sliding window on 200 SNPs at steps on 5 SNPs pruning the datasets to contain only one of 2 correlated SNPs with a r^2>0.5. To plot enrichment, we equally plotted observed P-values against expected under the null hypothesis (QQ plots). Enrichment and QQ plots were plotted overall for all 290 loci and separately for each of the 16 autoimmune diseases, however only for those diseases with more than 10 loci reported in the GWA catalogue. For extracted SNPs a False Discovery Rate corrected P-value < 0.05 was considered significant. Analysis were performed in Plink[7] and R project (3.0.1)[8].

Functional evaluation

Enrichment of SNPs falling in DHS sites:

DHS sites were downloaded from the ENCODE project[9] and from the Epigenomics Roadmap[10] selecting only cell types (or cell lines, herefrom "cell types") with duplicates, removing transformed cell types, and removing redundant cell lines, based on manual curation. Huh7 was an outlier in all analyses, and was removed. DHS sites were set to a fixed width of 150bp from center of region for all cell types. Allergy and Crohn's Disease SNPs were split in bins of increasing p-value cutoff, starting at 1 (including all SNPs and setting baseline for enrichment) and decreasing one decimal digitor each bin (1, 0.1, 0.001 etc). Each bin was overlapped with DHS regions using bedtools v2.19.0, and enrichment for each bin was calculated for each cell type as compared to p =1. For GWAS Catalogue SNPs, SNPs were selected for traits with > 30 reported associated SNPs. For identical SNPs for the same trait, the SNP with the lowest p-value was chosen. SNPs were then overlapped with DHS regions using bedtools v2.19.0, and ratio of overlapping SNPs was calculated. To filter out non-informative cell-types, only cell-types with the highest quartile of overlap ratios was included. For immune cell hierarchical clustering the manhattan distance of square root transformed ratios were used. For PCA, log10 transformed ratios were used.

Enrichment of SNPs falling in FANTOM enhancers:

FANTOM cell specific enhancers were downloaded from 'http://enhancer.binf.ku.dk/' and were set to a fixed width of 150bp from center of region for all cell types. The ratio between overlaps of all SNPs (p <= 1) and SNPs at p<= 1e-5 was calculated, and a p-value for this ratio calculated using a binomial test with the genomic overlap frequency as null frequency, calculated as the number of total enhancers per cell type times enhancer length (150bp), divided by the total number of base pairs shown to be bound by transcription factors in the human genome across cell types in the ENCODE project (231mb)[9]. FDR values were calculated adjusting p-values with the Benjamini-Hochberg method.

Data-driven Enrichment-Prioritized Integration for Complex Traits (DEPICT):

For details of this method please refer to Pers et al.[11] . DEPICT facilitates gene set enrichment analysis by testing whether genes in associated regions enrich for reconstituted versions of known pathways, gene sets, as well as protein complexes. The gene-set enrichment analyses in DEPICT contains three steps: first a scoring step; second a bias correcting step taking into account gene density that possibly could inflate results due to gene length and finally estimating experiment-wide FDR's.

Gene set reconstitution is accomplished by identifying genes that are co-expressed with genes in a given gene set based on a panel of 77,840 gene expression microarrays; genes that co-express with genes from a given gene set are likely to be part of that gene set.[12] In DEPICT, several types of gene sets were reconstituted: 5,984 protein complexes that were derived from 169,810 high-confidence experimentally-derived protein-protein interactions[13]; 2,473 phenotypic gene sets derived from 211,882 gene-phenotype pairs from the Mouse Genetics Initiative[14]; 737 Reactome database pathways[15]; 184 KEGG database pathways[16]; and 5,083 Gene Ontology database terms[17]. In addition, the DEPICT also facilitates tissue and cell type enrichment analysis, by testing whether genes in associated regions are highly expressed in any of 209 Medical Subject Heading annotations of 37,427 microarrays from the Affymetrix U133 Plus 2.0 Array platform. We used DEPICT to test enrichment in a total of 14,461 reconstituted gene sets and enrichment of 209 tissue and cell type annotations.

For the allergy meta-analysis, and Crohns disease, DEPICT was performed on all loci P<10e-5. For PCA GWAS catalogue data, DEPICT was performed on all traits with more than 30 reported associated SNPs. For identical SNPs for the same trait, the SNP with the lowest p-value was chosen.

Pathway analysis and visualization for allergy and Crohn's disease:

To account for difference in GWAS study sizes, a linear model was fitted between logged p-values of DEPICT results for allergy and Chron's disease, and the estimator was used to adjust the p-value thresholds for the largest study, Crohn's disease. The inflation estimate for Crohn's disease was 1.21. Shared pathways were set at $p_{allergy}$ and $p_{crohns\_adjusted} < 0.001$, allergy specific pathways were set at $p_{allergy} < 0.001$ and $p_{crohns\_adjusted} > 0.05$ and Crohn's disease specific pathways were set at $p_{allergy} > 0.05$ and $p_{crohns\_adjusted} < 0.001$.

DHS genomic location:

Genomic regions for 186 cell types, of which 14 cell types (CD14_Primary_Cells, CD19_Primary_Cells_Peripheral_UW, CD20, CD3_Primary_Cells_Cord_BI, CD3_Primary_Cells_Peripheral_UW, CD34Mobilized, CD56_Primary_Cells, Th0, Th1, Th17, Th2, Treg, GM12864, and Fetal_Thymus) were annotated as immune cells, were downloaded from the ROADMAP and ENCODE tracks (June 2014) in the UCSC Genome Browser and processed by bedtools, ensuring no redundancy between exons, introns, promoters (defined as 5000 bases upstream and 200 bases downstream of transcription start sites), and intergenic sites. DHS sites were overlapped with genomic regions, requiring 1bp of overlap. Enrichment of markers in DHS regions was calculated for GWAS catalog traits with 30 or more reported variants, and was normalized by trait SNP count and cell type specific DNAse sequence lenghts.

The uneven distribution of cell types within the Roadmap ENCODE dataset could possibly contribute to the separation of immune-mediated diseases from other diseases. However, repeated iterative removal of ¼ of cell types continuously produced a statistical significant separation of autoimmune diseases, allergy and asthma vs. other traits (results not shown), hence supporting the finding of common SNPs and cell types to congregate in allergy and autoimmune diseases. In addition, hierarchical clustering was performed on the full ENCODE Roadmap set to investigate if this clustering was facilitated by similar DHS-profiles in different immune cells, basically representing a single "immune-system-footprint", but this was not the case, as different immune cell types also separated internally, comparable to other non-immune cell types (**Supplementary Figure 18**).

A further cluster analysis of the cell-type specific genomic DHS location in all cells (intronic, exonic, intergenic, promotor) was performed revealing that the DHS sites in immune cells tend to fall within promotor and exonic regions (**Supplementary Figure 19**).

Transcription factor binding sites:

Transcription factor binding sites for 161 transcription sites were downloaded in BED format from ENCODE for the hg19 build, and were intersected with 28 independent shared loci, expanded to included markers with r2>= 0.5 in the 1000g CEU panel, using BedTools. Enrichment and one-sided p-values were calculated in relation to an empirical null distribution of loci overlap for each TF, generated by 10,000 random permutations of random genomic loci with the same length characteristics as the 28 LD-expanded shared loci**.** Random locus LD-structure was assumed to have no effect on TF-binding probability as a function of locus length.

Acknowledgements and funding

23andMe: We thank the customers of 23andMe who answered surveys, as well as the employees of 23andMe, who together made the 23andMe research possible. This work was supported in part by the National Heart, Lung, and Blood Institute of the US National Institutes of Health under grant 1R43HL115873-01.

AAGC: QIMR: We thank the twins and their families for their participation; Dixie Statham, Ann Eldridge, Marlene Grace, Kerrie McAloney (sample collection); Lisa Bowdler, Steven Crooks (DNA processing); David Smyth, Harry Beeby, Daniel Park (IT support). Busselton: The Busselton Health Study acknowledges the numerous Busselton community volunteers who assisted with data collection and the study participants from the Shire of Busselton.

The NHMRC (including grants 613627 and 1036550), Asthma Foundations in Tasmania, Queensland and Victoria, The Clifford Craig Trust in Northern Tasmania, Lew Carty Foundation, Royal Hobart Research Foundation and the University of Melbourne, Cooperative Research Centre for Asthma, New South Wales Department of Health, Children's Hospital Westmead, University of Sydney. Contributions of goods and services were made to the CAPS study by Allergopharma Joachim Ganzer KG Germany, John Sands Australia, Hasbro, Toll refrigerated, AstraZeneca Australia, and Nu-Mega Ingredients Pty Ltd. Goods were provided at reduced cost to the CAPS study by Auspharm, Allersearch and Goodman Fielder Foods. MCM, SCD and MAB are supported by the NHMRC Fellowship Scheme. The Busselton Health Study acknowledges the generous support for the 1994/5 follow-up study from Healthway, Western Australia.

ALSPAC: We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and

nurses. The UK Medical Research Council and the Wellcome Trust (Grant ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. This publication is the work of the authors Nicholas Timpson and John Henderson will serve as guarantors for the contents of this paper. This research was specifically funded by The MRC Centre for Causal Analyses in Translational Medicine (grant G0600705) provided funding for N Timpson.

GWAS data was generated by Sample Logistics and Genotyping Facilities at the Wellcome Trust Sanger Institute and LabCorp (Laboratory Corportation of America) using support from 23andMe.

B58C: We acknowledge use of phenotype and genotype data from the British 1958 Birth Cohort DNA collection, funded by the Medical Research Council grant G0000934 and the Wellcome Trust grant 068545/Z/02. (http://www.b58cgene.sgul.ac.uk/). Genotyping for the B58C-WTCCC subset was funded by the Wellcome Trust grant 076113/B/04/Z. The B58C-T1DGC genotyping utilized resources provided by the Type 1 Diabetes Genetics Consortium, a collaborative clinical study sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Allergy and Infectious Diseases (NIAID), National Human Genome Research Institute (NHGRI), National Institute of Child Health and Human Development (NICHD), and Juvenile Diabetes Research Foundation International (JDRF) and supported by U01 DK062418. B58C-T1DGC GWAS data were deposited by the Diabetes and Inflammation Laboratory, Cambridge Institute for Medical Research (CIMR), University of Cambridge, which is funded by Juvenile Diabetes Research Foundation International, the Wellcome Trust and the National Institute for Health Research Cambridge Biomedical Research Centre; the CIMR is in receipt of a Wellcome Trust Strategic Award (079895). The B58C-GABRIEL genotyping was supported by a contract from the European Commission Framework Programme 6 (018996) and grants from the French Ministry of Research.

COPSAC2000: We thank all the families participating in the COPSAC cohort for their effort and commitment; and the whole COPSAC study team including computer and laboratory technicians,

research scientists, managers, receptionists, and nurses. We also wish to thank Bjarne Kristensen and Inger Pedersen (Phadia Denmark) for their dedicated work with the IgE-analyses.

COPSAC is funded by: the Lundbeck Foundation, the Danish Council for Strategic Research, ,the Pharmacy Foundation, the Danish Agency for Science, Technology and Innovation, the EU Seventh Framework Programme, Ronald McDonald House Charities, the Global Excellence in Health award Programme, the Danish Medical Research Council, the Director K. GAD and family Foundation, the A. P. Møller og Hustru Chastine Mc-Kinney Møller General Purpose Foundation, the Aage Bang Foundation, the Health Insurance Foundation, the East Danish Medical Research Council, the Copenhagen City Council Research Foundation, the Kai and Gunhild Lange Foundation, the Dagmar Marshall Foundation, the Ville Heise legacy, the Region of Copenhagen, the Ib Henriksen foundation, the Birgit and Svend Pock-Steen foundation, the Danish Ministry of the Interior and Health's Research Centre for Environmental Health, the Gerda and Aage Hensch foundation, the Rosalie Petersens Foundation, the Hans and Nora Buchard Foundation, the Gangsted Foundation, the Danish Medical Association, Asthma-Allergy Denmark, the Danish Otolaryngology Association, the Oda Pedersen legacy, the Højmosegaard Legacy, the A. P. Møller og Hustru Chastine Mc-Kinney Møller Foundation for the advancement of Medical Knowledge, the Jacob and Olga Madsen Foundation, the Aase and Einar Danielsen Foundation, and Queen Louise's Children's' Hospital Research Foundation.

The IgE analyses were sponsored by Phadia ApS.

The funding agencies did not have any role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

LISA/GINI: The authors thank all the families for their participation in the GINIplus and LISAplus studies. Furthermore, we thank all members of the GINIplus and LISAplus Study Groups for their excellent work. The LISAplus Study Group consists of the following: The study team wishes to acknowledge the following: Helmholtz Zentrum Muenchen - German Research Center for Environment and Health, Institute of Epidemiology I, Neuherberg (Heinrich J, Wichmann HE,

Sausenthaler S, Chen C-M); University of Leipzig, Department of Pediatrics (Borte M), Department of Environmental Medicine and Hygiene (Herbarth O); Department of Pediatrics, Marien-Hospital, Wesel (von Berg A); Bad Honnef (Schaaf B); UFZ-Centre for Environmental Research Leipzig-Halle, Department of Environmental Immunology (Lehmann I); IUF – Leibniz Research Institute for Environmental Medicine, Düsseldorf (Krämer U); Department of Pediatrics, Technical University, Munich (Bauer CP, Hoffman U); Centre for Allergy and Environment, Technical University, Munich (Behrendt H).

The GINIplus Study Group consists of the following: The study team wishes to acknowledge the following: Helmholtz Zentrum Muenchen - German Research Center for Environmental Health, Institute of Epidemiology I, Munich (Heinrich J, Wichmann HE, Sausenthaler S, Chen C-M, Thiering E, Tiesler C, Standl M, Schnappinger M, Rzehak P); Department of Pediatrics, Marien-Hospital, Wesel (Berdel D, von Berg A, Beckmann C, Groß I); Department of Pediatrics, Ludwig Maximilians University, Munich (Koletzko S, Reinhardt D, Krauss-Etschmann S); Department of Pediatrics, Technical University, Munich (Bauer CP, Brockow I, Grübl A, Hoffmann U); IUF – Leibniz Research Institute for Environmental Medicine, Düsseldorf (Krämer U, Link E, Cramer C); Centre for Allergy and Environment, Technical University, Munich (Behrendt H).

Personal and financial support by the Munich Center of Health Sciences (MCHEALTH) as part of the Ludwig-Maximilians University Munich LMU innovative is gratefully acknowledged.

Manchester Asthma and Allergy Study (MAAS): We would like to thank the children and their parents for their continued support and enthusiasm. We greatly appreciate the commitment they have given to the project. We would also like to acknowledge the hard work and dedication of the study team (post-doctoral scientists, research fellows, nurses, physiologists, technicians and clerical staff).

MAAS was supported by the Asthma UK Grants No 301 (1995-1998), No 362 (1998-2001), No 01/012 (2001-2004), No 04/014 (2004-2007) and The Moulton Charitable Foundation (2004-current); age 11 years clinical follow-up is funded by the Medical Research Council (MRC) Grant G0601361.

NFBC 1966: We thank Professor Paula Rantakallio (launch of NFBC 1966 and 1986), Ms Outi Tornwall and Ms Minttu Jussila (DNA biobanking).

Financial support was received from the Academy of Finland (project grants 104781, 120315 and Center of Excellence in Complex Disease Genetics), University Hospital Oulu, Biocenter, University of Oulu, Finland, the European Commission (EURO-BLCS, Framework 5 award QLG1-CT-2000-01643), NHLBI grant 5R01HL087679-02 through the STAMPEED program (1RL1MH083268-01), NIH/NIMH (5R01MH63706:02), ENGAGE project and grant agreement HEALTH-F4-2007-201413, and the Medical Research Council (G0500539, PrevMetSyn/SALVE). The DNA extractions, sample quality controls, biobank up-keeping and aliquotting was performed in the National Public Health Institute, Biomedicum Helsinki, Finland and supported financially by the Academy of Finland and Biocentrum Helsinki. A. Couto Alves acknowledges the European Commission, Framework 7, grant number 223367. Jess L Buxton acknowledges the Wellcome Trust fellowship grant, number WT088431MA.

RAINE: The authors are grateful to the Raine Study participants, their families, and to the Raine Study research staff for cohort coordination and data collection. The authors gratefully acknowledge the assistance of the Western Australian DNA Bank (National Health and Medical Research Council of Australia National Enabling Facility). The following Institutions provide funding for Core Management of the Raine Study: The University of Western Australia (UWA), Raine Medical Research Foundation, UWA Faculty of Medicine, Dentistry and Health Sciences, The Telethon Institute for Child Health Research, Curtin University, Edith Cowan University and Women and Infants Research Foundation. This study was supported by project grants from the National Health and Medical Research Council of

Australia (Grant ID 403981 and ID 003209; http://www.nhmrc.gov.au/) and the Canadian Institutes of Health Research (Grant ID MOP-82893; http://www.cihr-irsc.gc.ca/e/193.html).

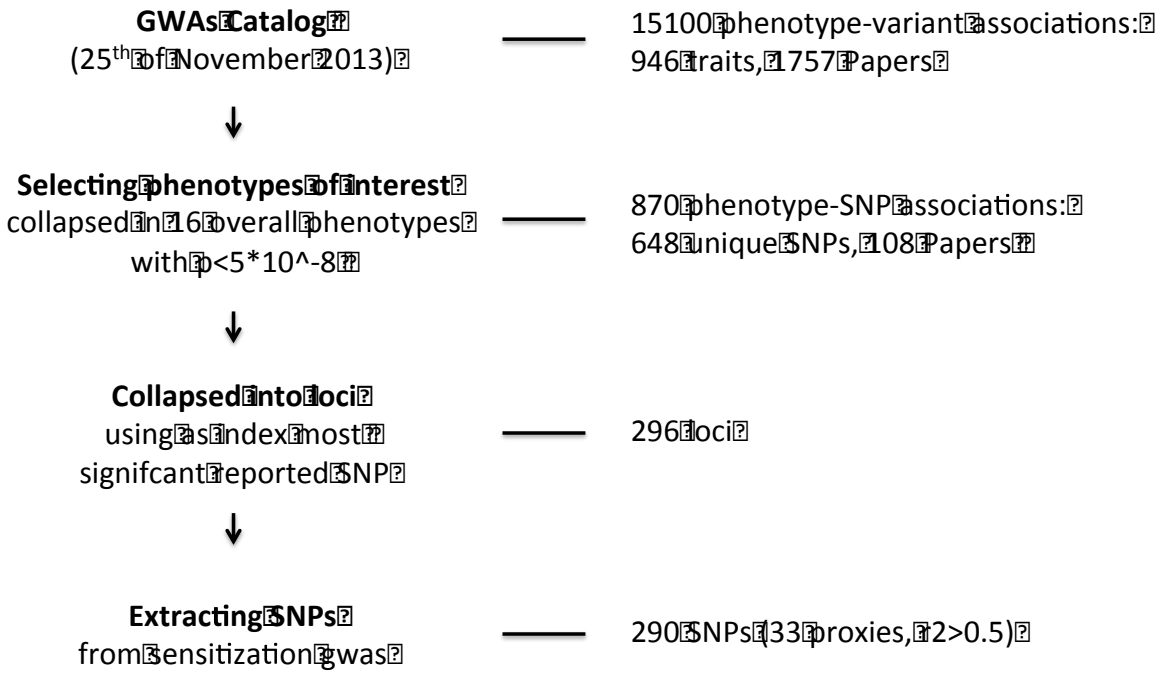
PIAMA: The PIAMA study would like to thank all participants, and co-investigators of the study.

The PIAMA study is supported by the Dutch Asthma Foundation (grant 3.4.01.26, 3.2.06.022, 3.4.09.081 and 3.2.10.085CO), the ZonMw (a Dutch organization for health research and development; grant 912-03-031), and the ministry of the environment.

Genome-wide genotyping was funded by the European Commission as part of GABRIEL (A multidisciplinary study to identify the genetic and environmental causes of asthma in the European Community) contract number 018996 under the Integrated Program LSH-2004-1.2.5-1 Post genomic approaches to understand the molecular basis of asthma aiming at a preventive or therapeutic control.

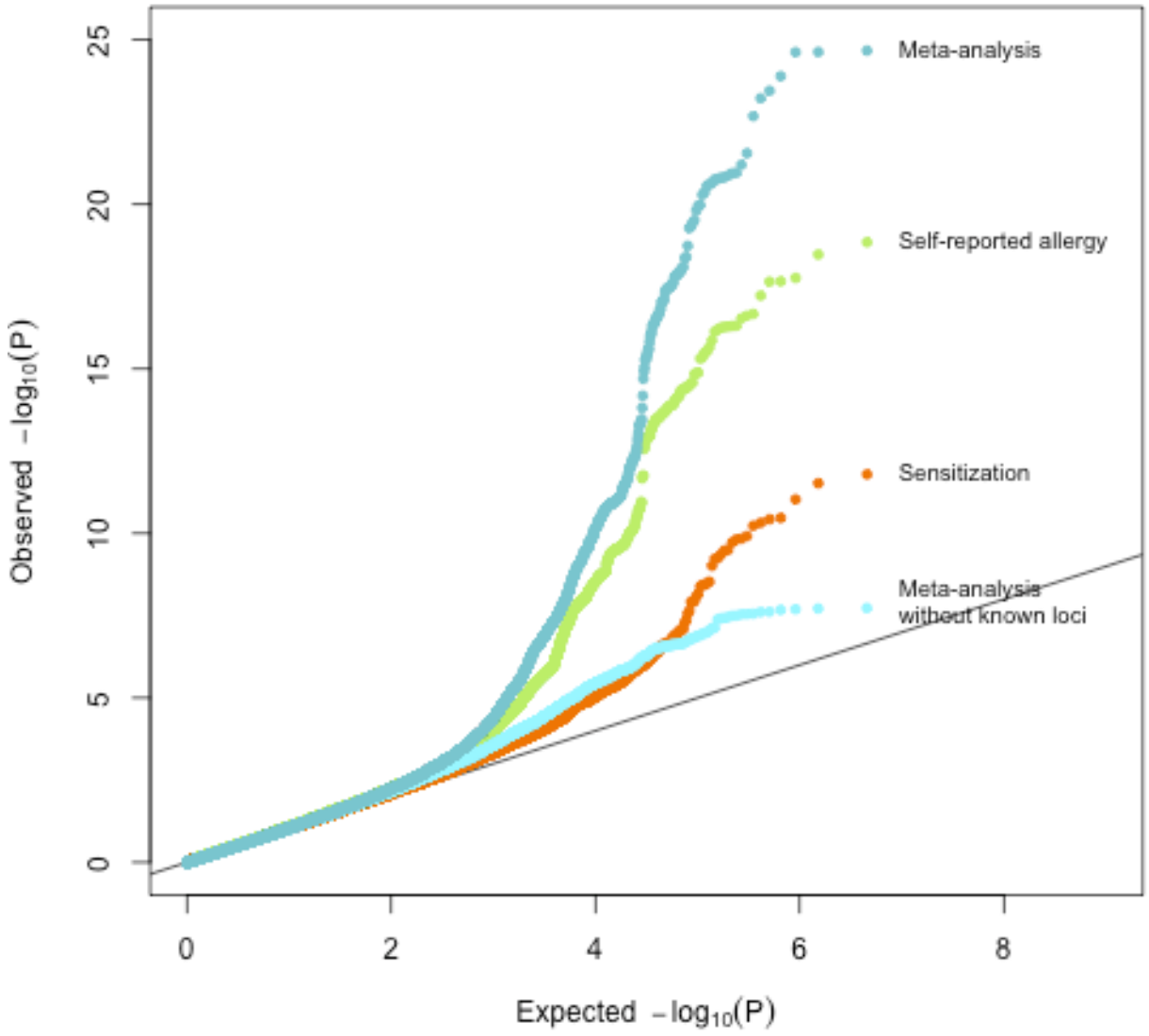
Supplementary Figure 1

Flowchart of the selected known autoimmune disease associated SNPs/loci for lookup in the allergy GWAS identified within the GWA's catalog (accessed 25th of November 2013). Please also see methods section here in the supplement. A detailed description for each step in the flow chart:
1) The GWA's catalog[5] were accessed 25th of November 2013.
2) All autoimmune diseases and associations to SNPs were selected with p<5*10^-8. The chosen traits were collapsed into 16 overall autoimmune disease categories (see supplemental table 1)
3) We collapsed close SNPs into loci (+/-250kb)[6] and used for each locus only the SNP with lowest reported P as index SNP and as representative for the specific locus.
4) For several of the SNPs we had to use a proxy SNP as the index SNP were not present within the allergy GWAS. Proxy SNPs were chosen on highest r2 to index SNP and if two or more proxies had the same r2 the SNP closest in physical distance to the index SNP were chosen. In total 290 SNPs were available for look up/extraction within the allergy GWAS.



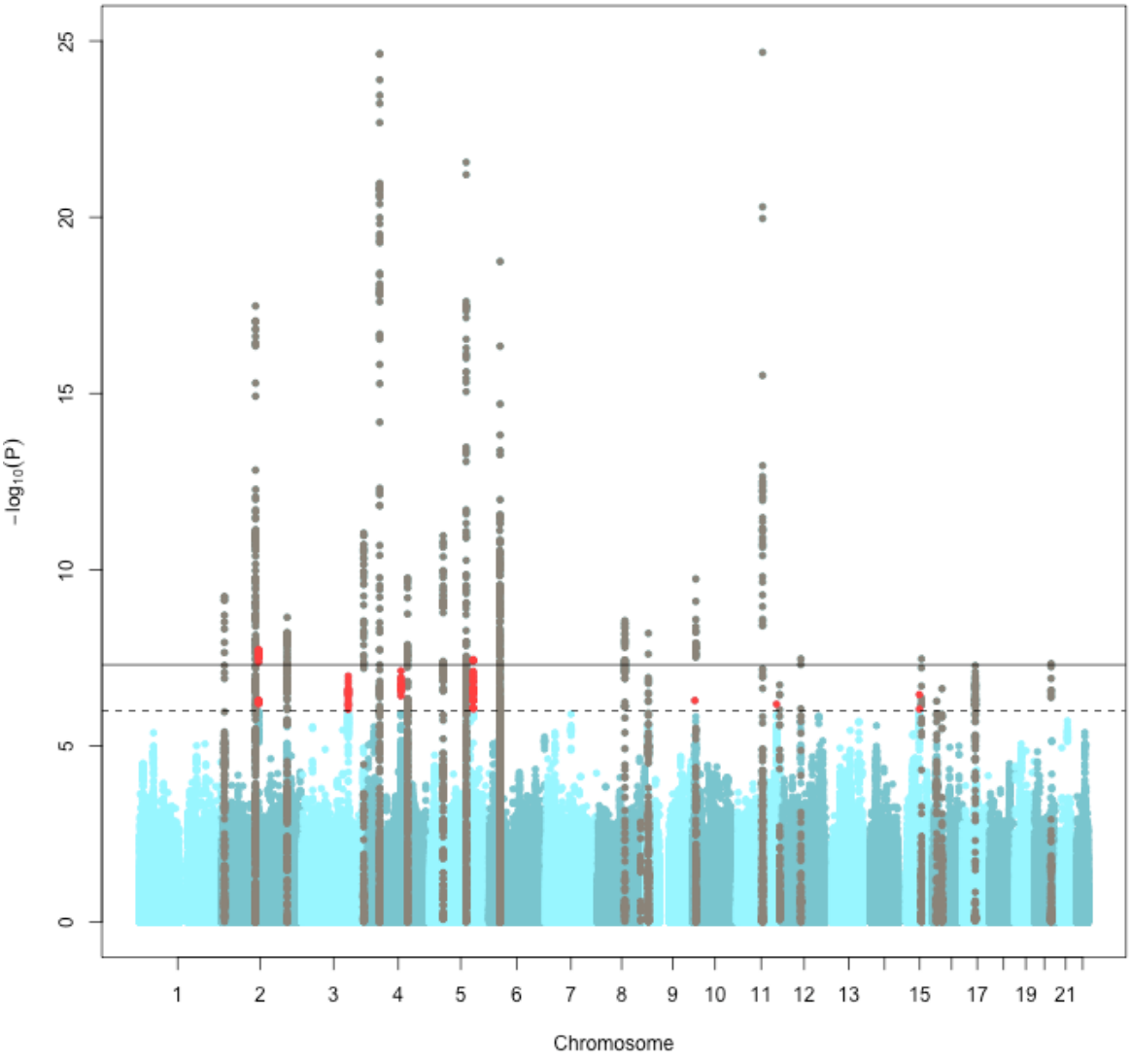
Supplementary Figure 2

QQ plot of the of the meta-analysed 2,284,215 SNPs and association to 1) Sensitization[2] 2) Self-reported allergy[3] 3) These two data-sets meta-analysed and 4) Without reported known loci



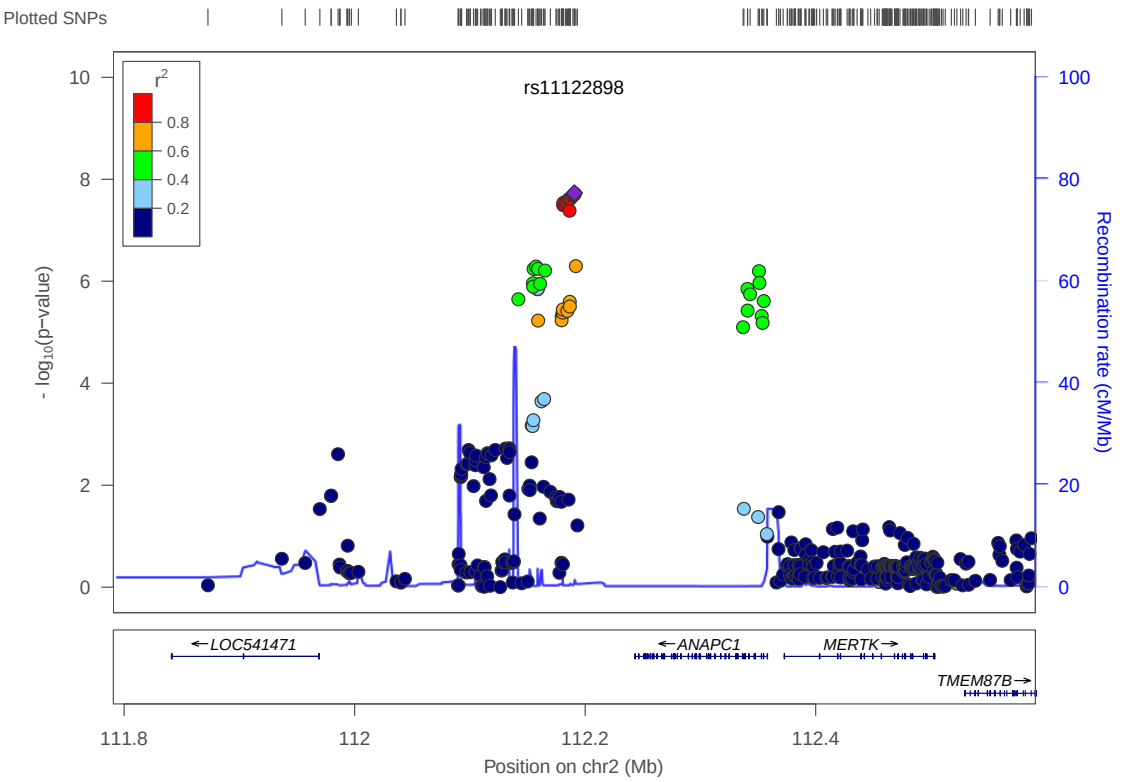
Supplementary Figure 3

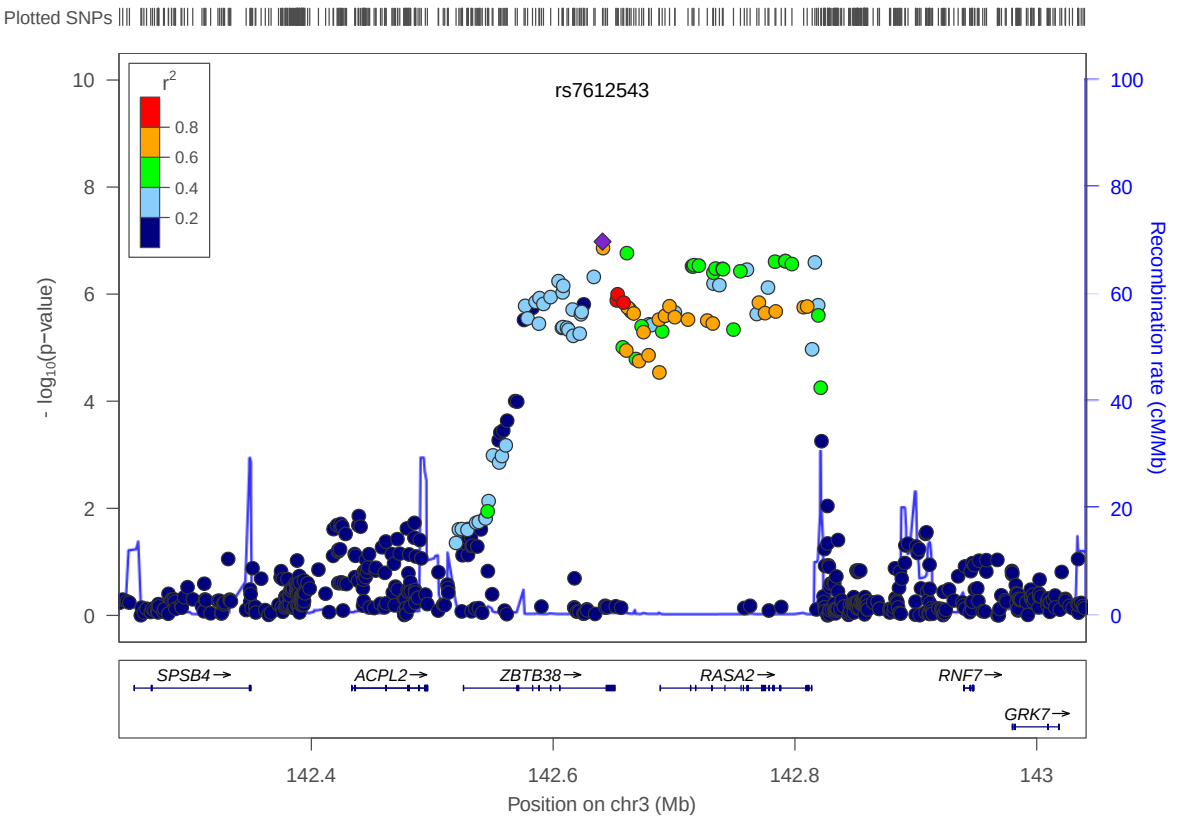
Manhattan plot of the of the meta-analysed 2,284,215 SNPs and association to allergy. Red dots indicate novel loci not described in the discovery papers (grey)[2,3], with p<5*10e-6. Dashed line: 10e-6. Solid line: 5e-8.



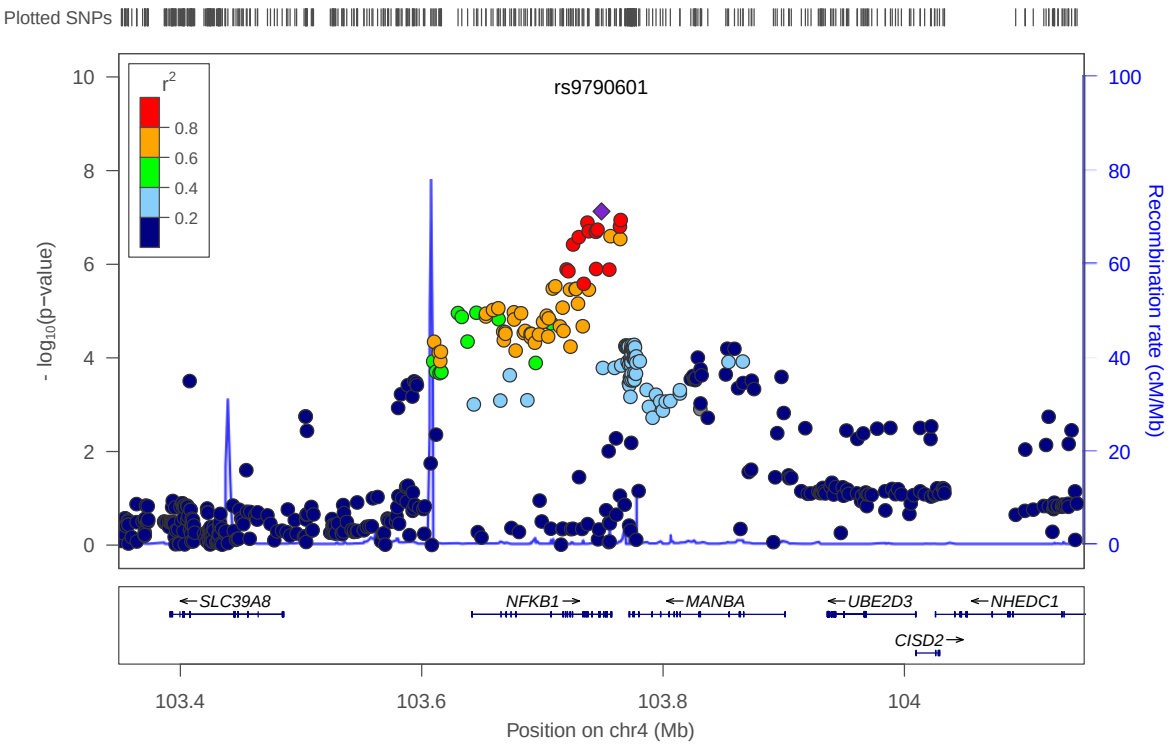
Supplementary Figure 4

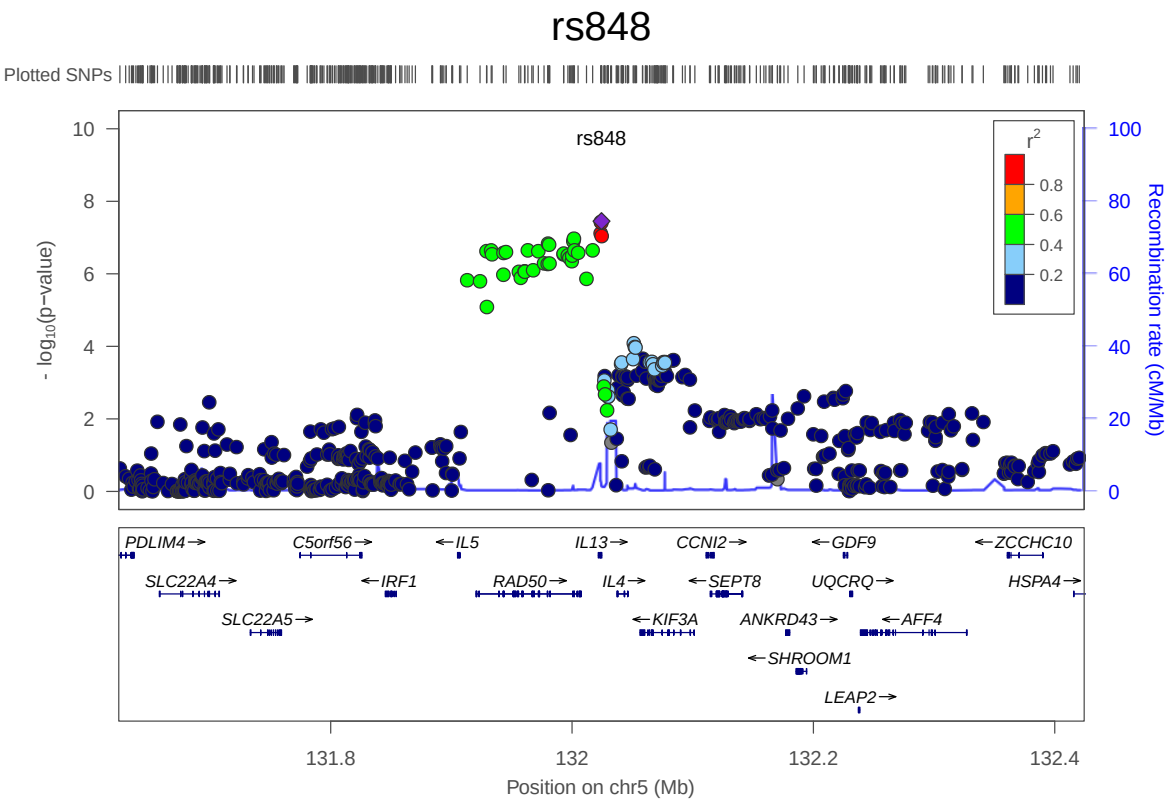
LocusZoom plots of the suggestive novel loci from the allergy meta-analyses



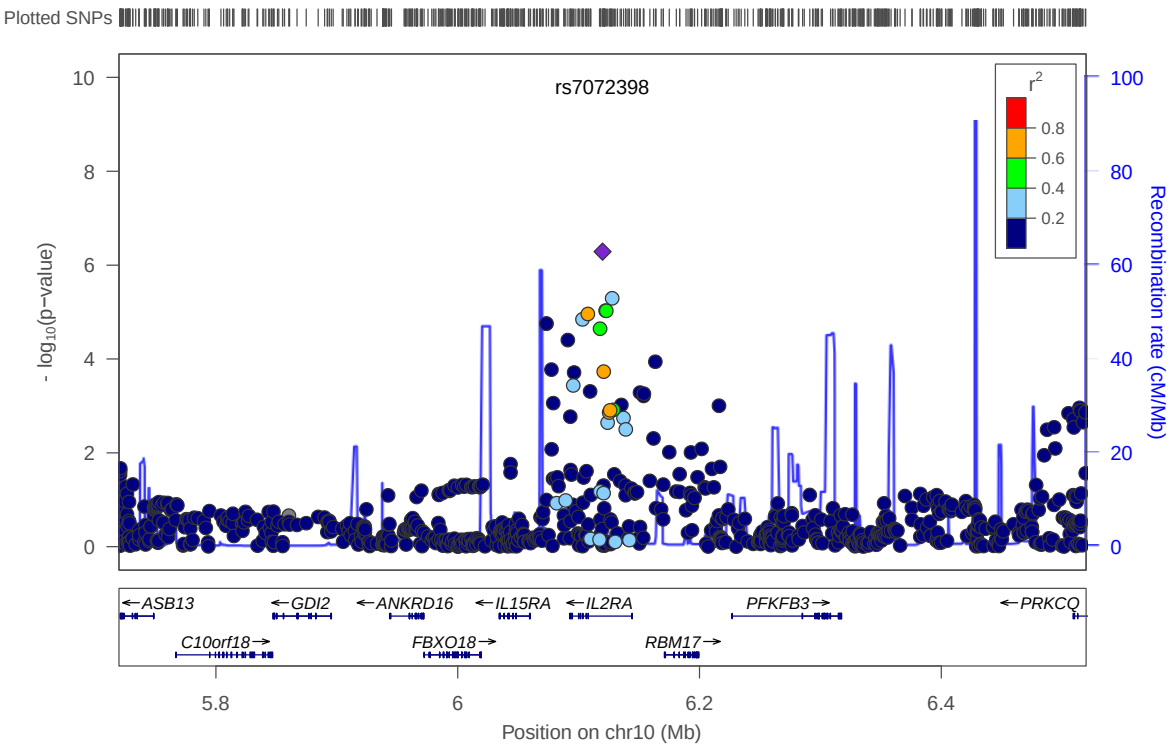

Plotted SNPs
rs7612543
r2
0.8
0.6
0.4
0.2
- log10(p-value)
Recombination rate (cM/Mb)
SPSB4
ACPL2
ZBTB38
RASA2
RNF7
GRK7
142.4
142.6
142.8
143
Position on chr3 (Mb)

rs9790601

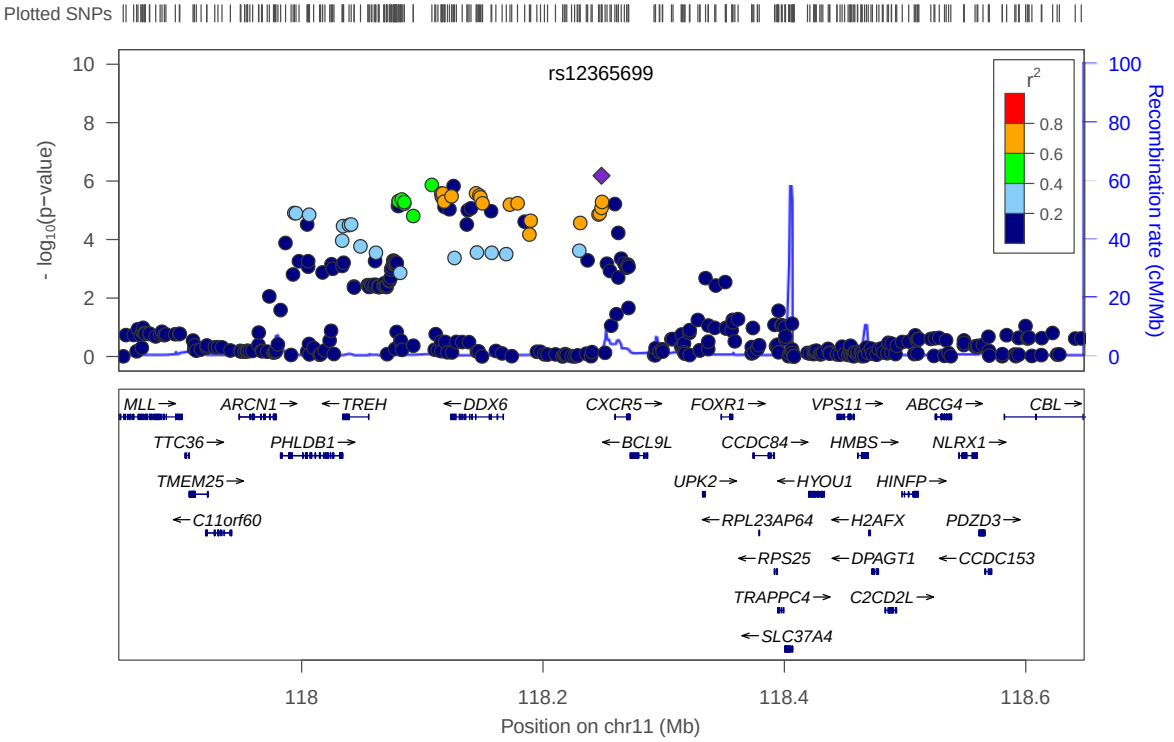

Plotted SNPs
rs9790601
r²
0.8
0.6
0.4
0.2
- log10(p−value)
Recombination rate (cM/Mb)
SLC39A8
NFKB1
MANBA
UBE2D3
NHEDC1
CISD2
103.4
103.6
103.8
104
Position on chr4 (Mb)


rs848
Plotted SNPs
rs848
r2
0.8
0.6
0.4
0.2
- log10(p−value)
Recombination rate (cM/Mb)
PDLIM4
C5orf56
IL5
IL13
CCNI2
GDF9
ZCCHC10
SLC22A4
IRF1
RAD50
IL4
SEPT8
UQCRQ
HSPA4
SLC22A5
KIF3A
ANKRD43
AFF4
SHROOM1
LEAP2
131.8
132
132.2
132.4
Position on chr5 (Mb)

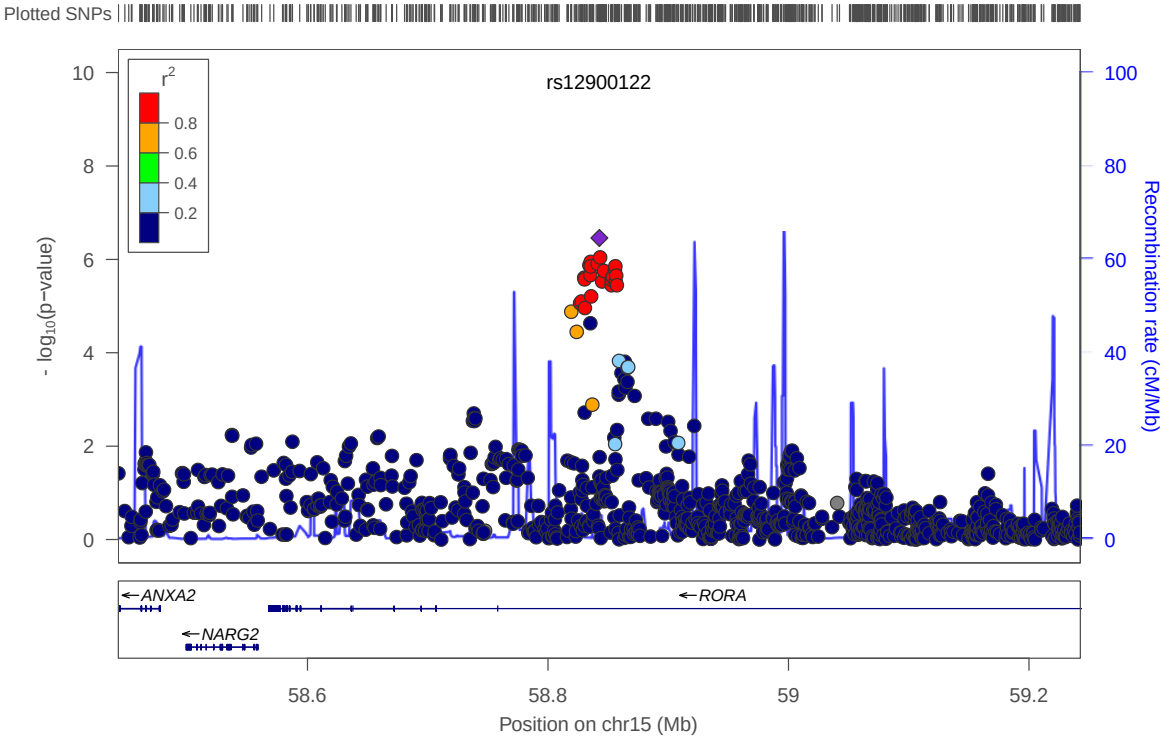
rs7072398


Plotted SNPs
rs7072398
r²
0.8
0.6
0.4
0.2
- log10(p−value)
Recombination rate (cM/Mb)
ASB13
GDI2
ANKRD16
IL15RA
IL2RA
PFKFB3
PRKCQ
C10orf18
FBXO18
RBM17
Position on chr10 (Mb)

rs12365699

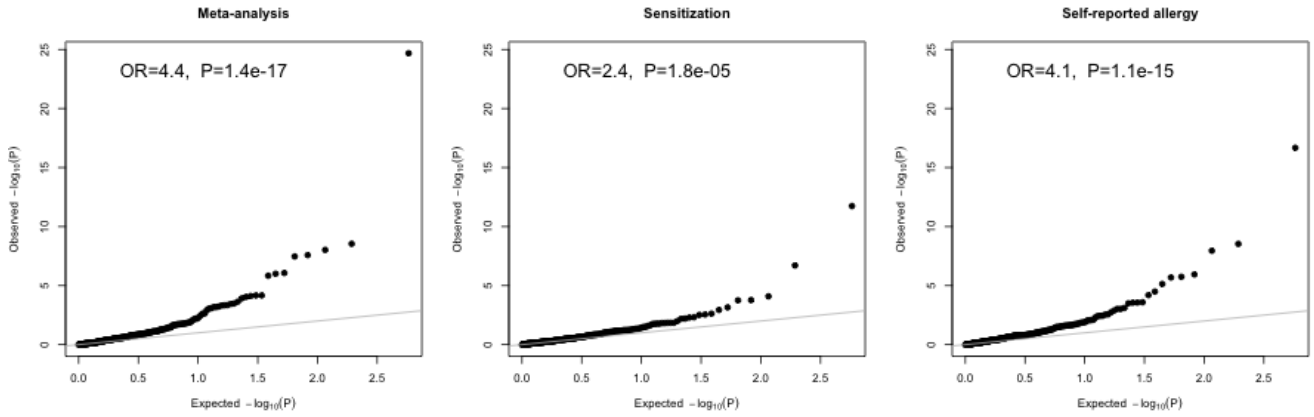

Plotted SNPs
rs12365699
r²
0.8
0.6
0.4
0.2
- log10(p−value)
Recombination rate (cM/Mb)
MLL
ARCN1
TREH
DDX6
CXCR5
FOXR1
VPS11
ABCG4
CBL
TTC36
PHLDB1
BCL9L
CCDC84
HMBS
NLRX1
TMEM25
UPK2
HYOU1
HINFP
C11orf60
RPL23AP64
H2AFX
PDZD3
RPS25
DPAGT1
CCDC153
TRAPPC4
C2CD2L
SLC37A4
118
118.2
118.4
118.6
Position on chr11 (Mb)

rs12900122


Plotted SNPs
rs12900122
r²
0.8
0.6
0.4
0.2
- log10(p−value)
Recombination rate (cM/Mb)
ANXA2
RORA
NARG2
58.6
58.8
59
59.2
Position on chr15 (Mb)

Supplementary Figure 5

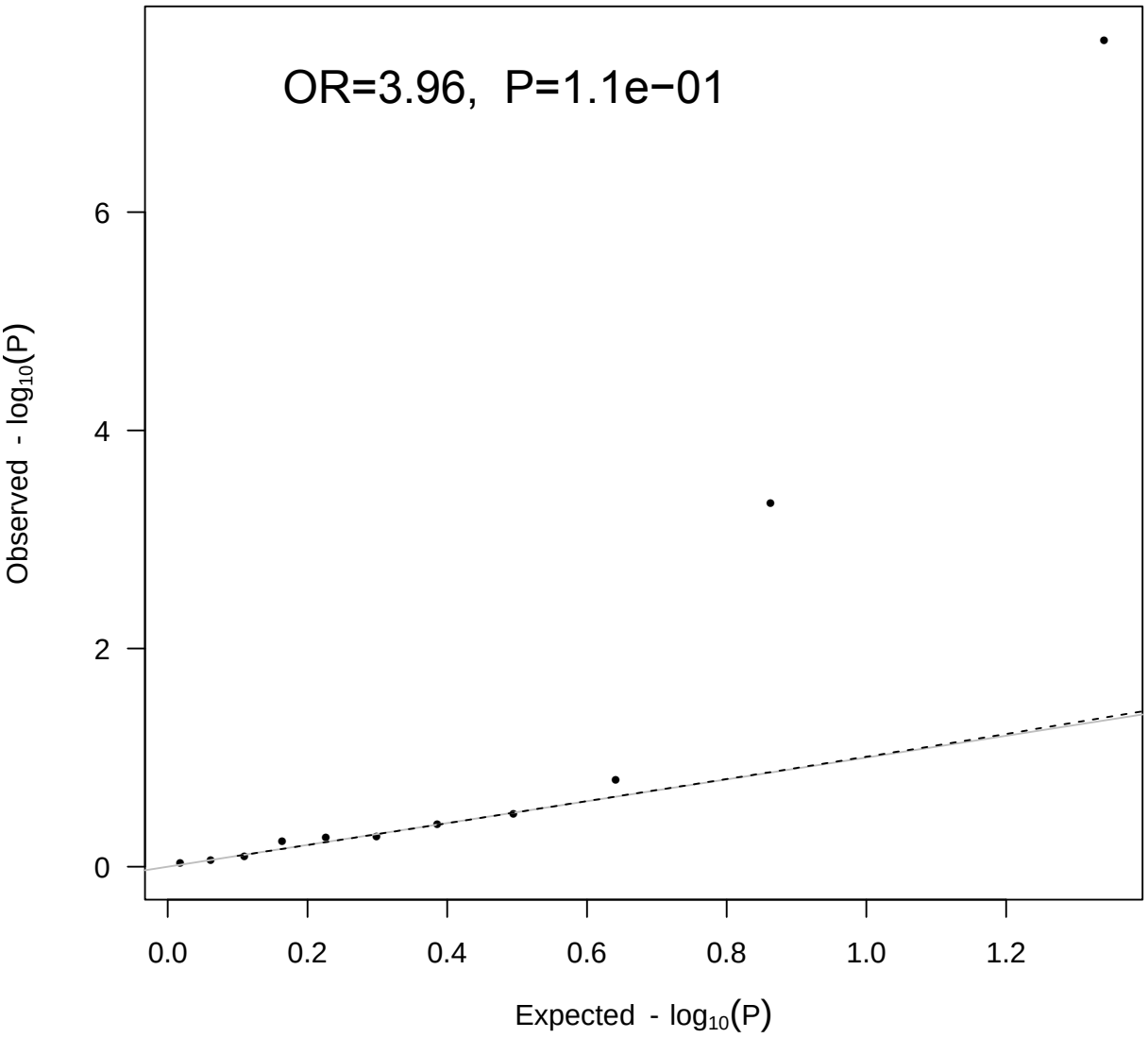
QQ plots of the autoimmune disease associated loci within the combined allergy meta-analysis as well as allergic sensitization and self-reported allergy separately. The numbers in the figures show enrichment Odds Ratio and P-value for enrichment.



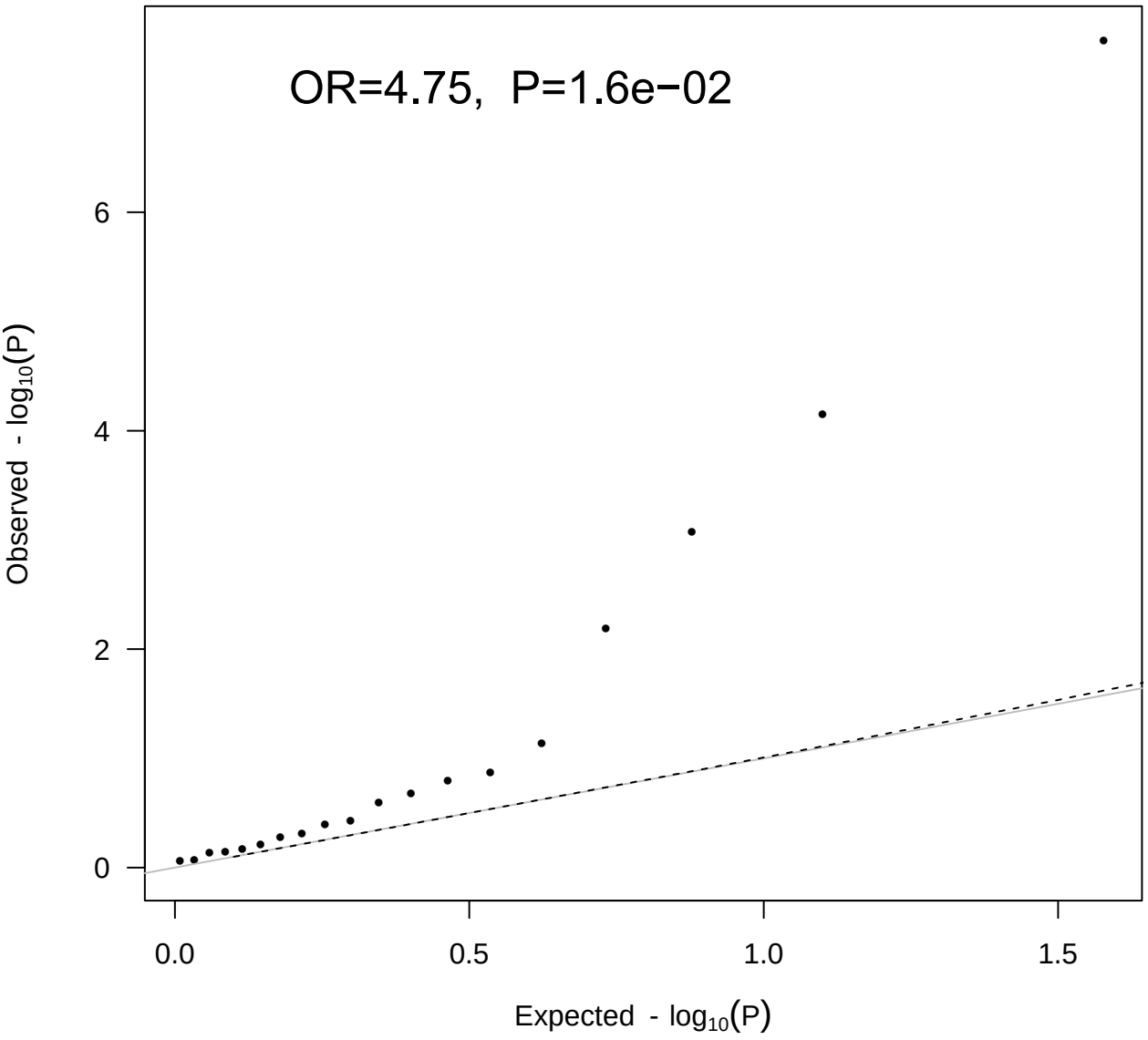
Supplementary Figure 6

Separate QQ plots of the autoimmune disease associated loci within the allergy meta-analysis with printed calculated enrichment Odds Ratio and P-value for enrichment. Only plotted for autoimmune diseases with at least 10 loci associated. Solid line reflects the P-value distribution under the null while the dashed is the distribution of all SNPs from the allergy meta-analysis.

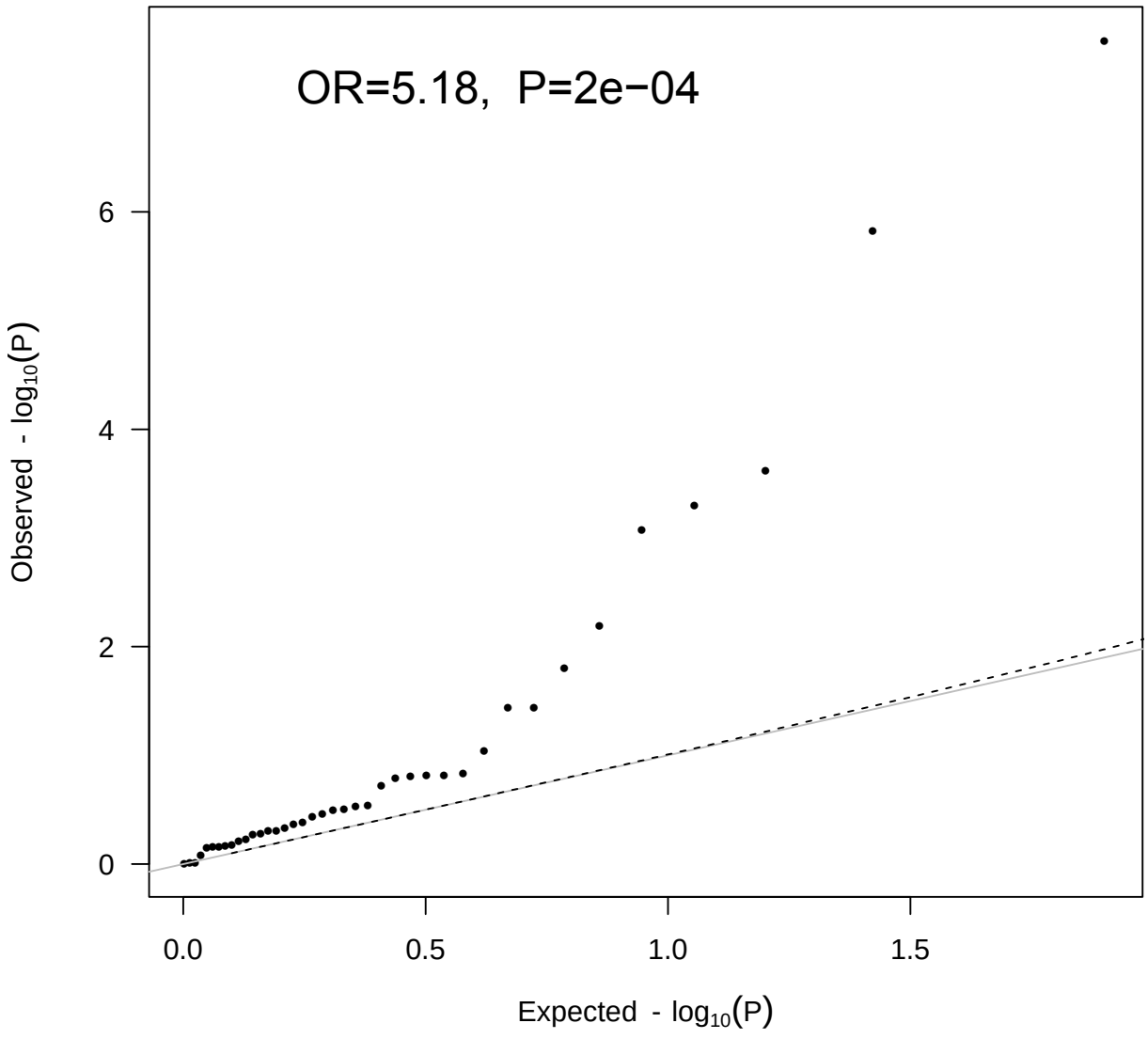
Ankylosing Spondylitis:



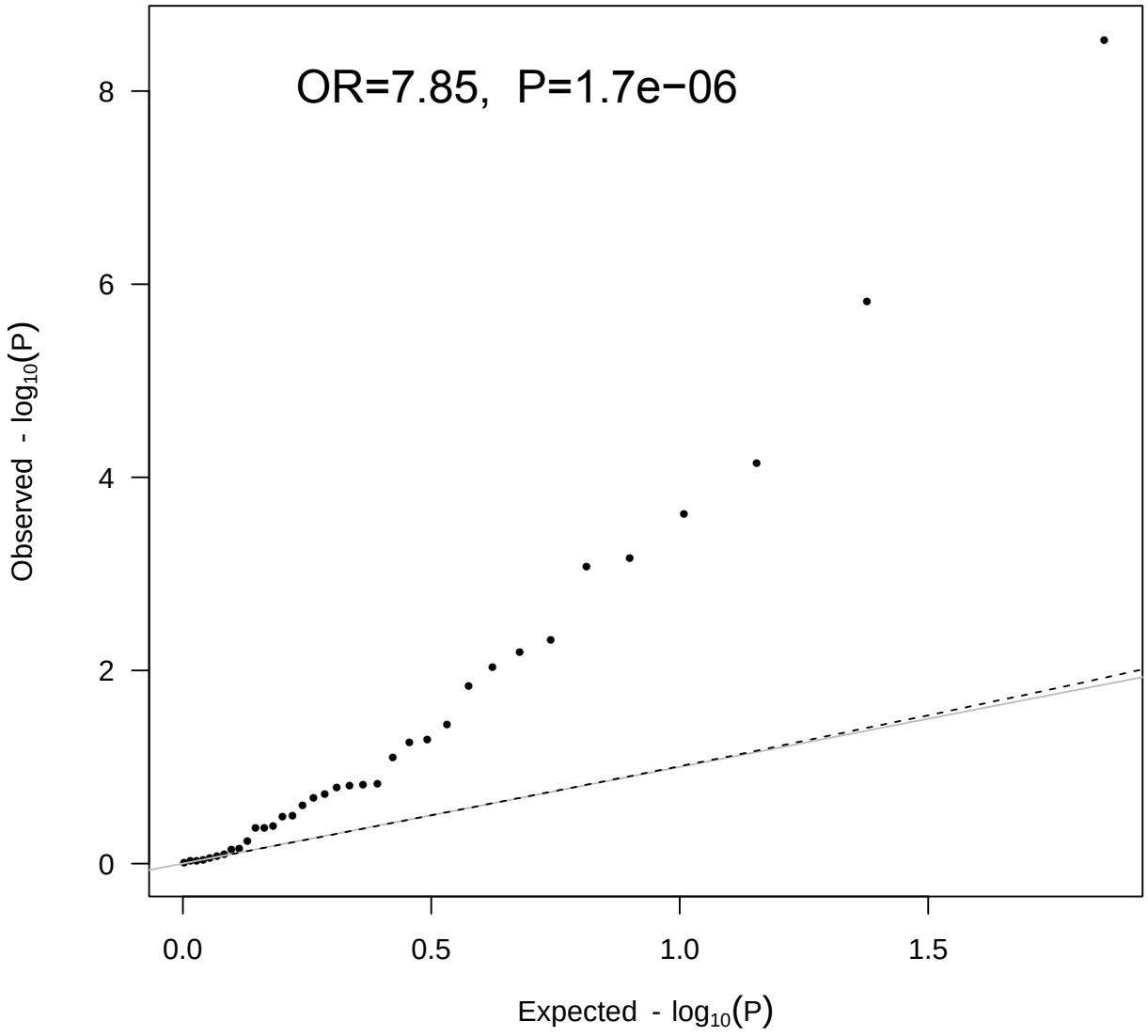
Psoriasis:


OR=4.75, P=1.6e−02
Observed - log10(P)
Expected - log10(P)
0
2
4
6
0.0
0.5
1.0
1.5

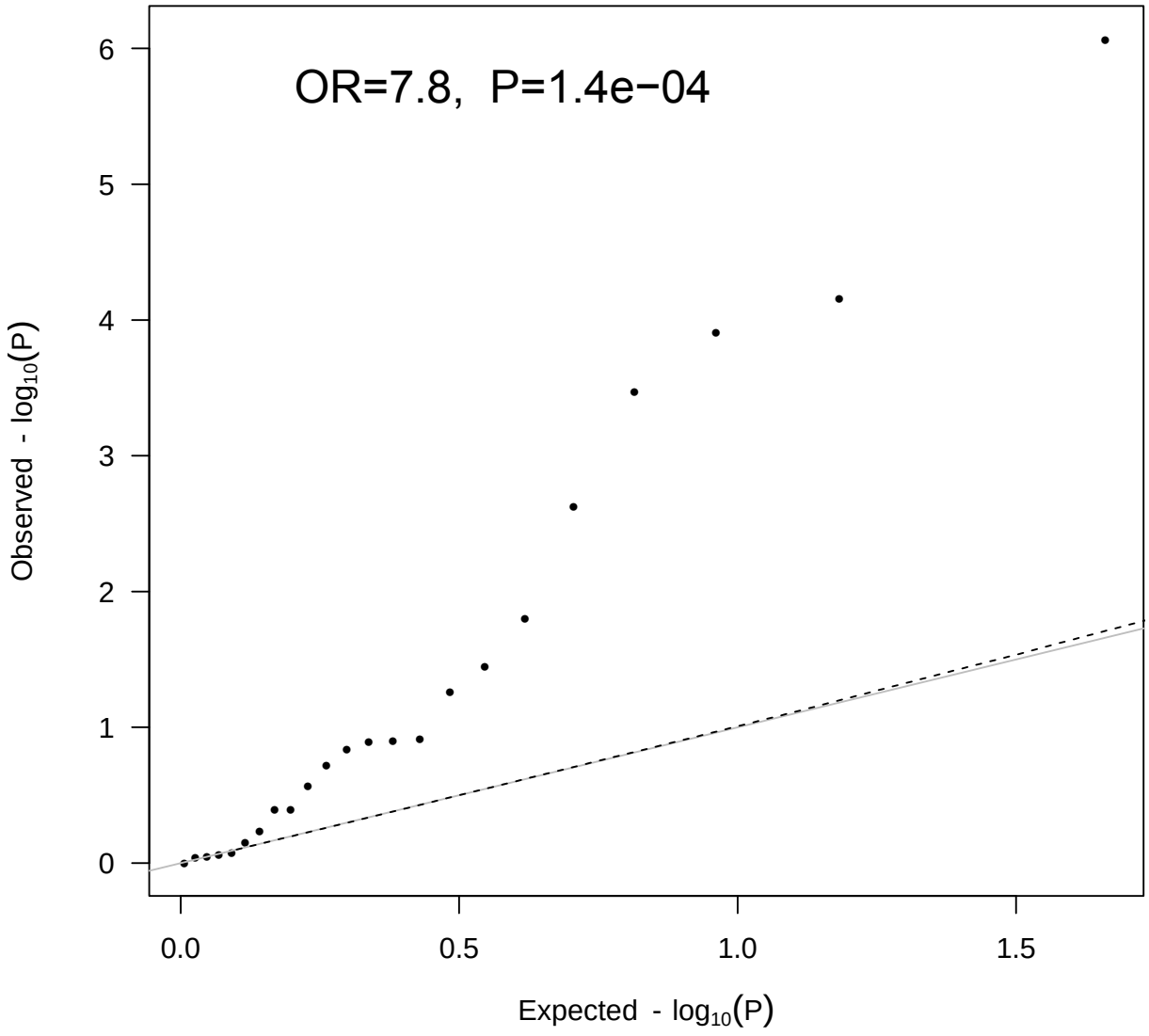
Arthirits:


OR=5.18, P=2e−04
Observed - log10(P)
Expected - log10(P)
0
2
4
6
0.0
0.5
1.0
1.5

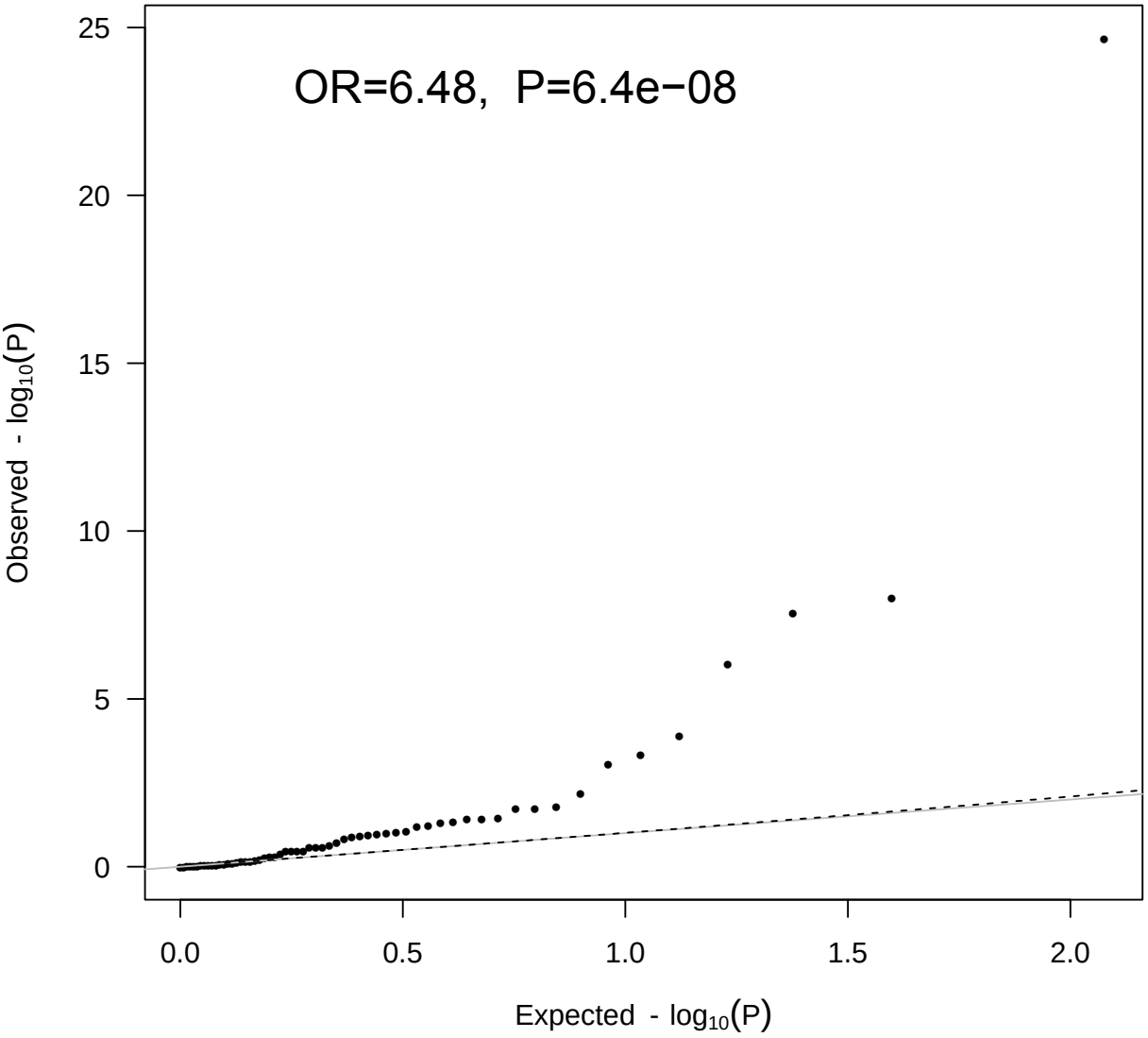
Celiac Disease:



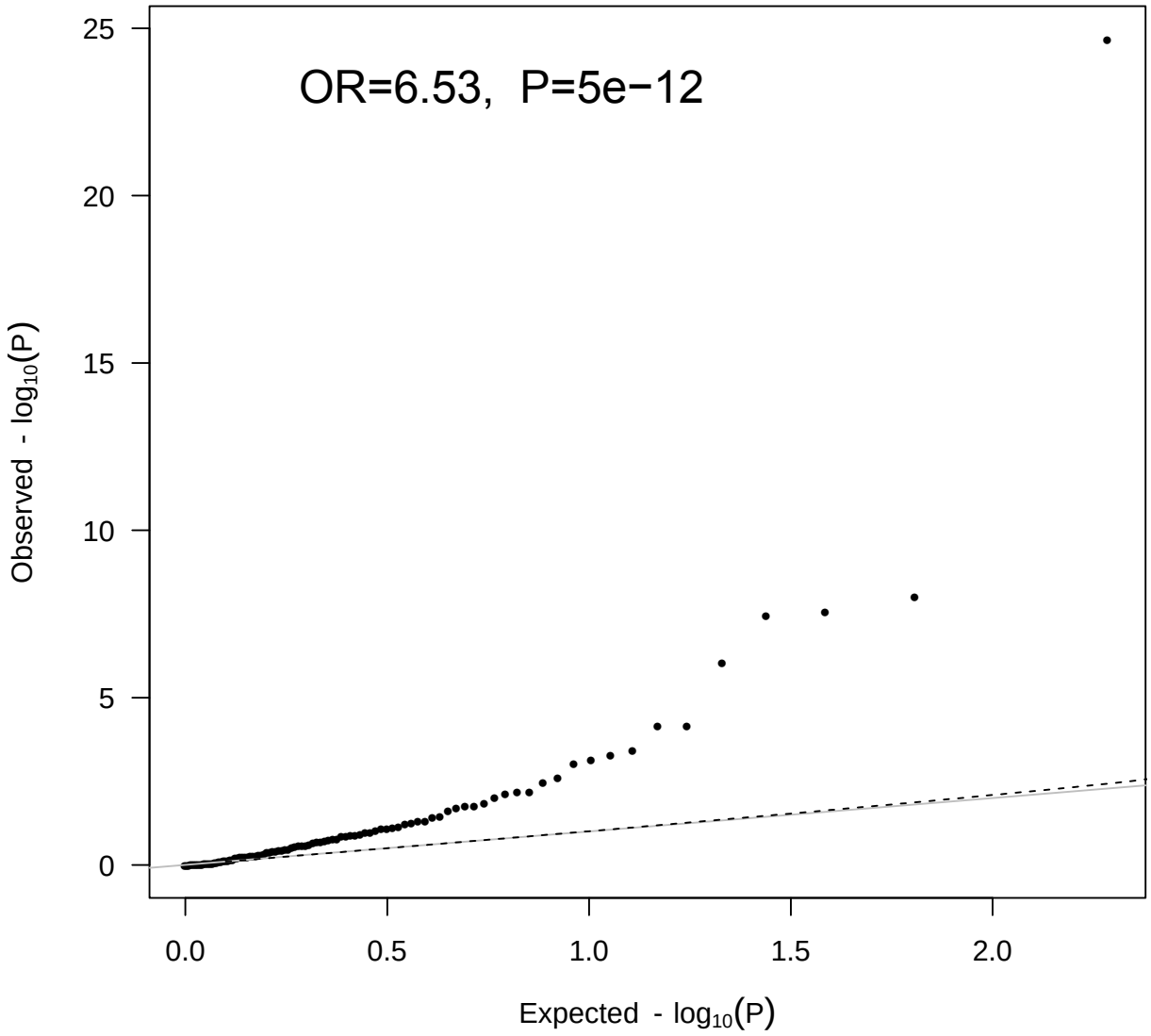
Primary Biliary Cirrhosis:



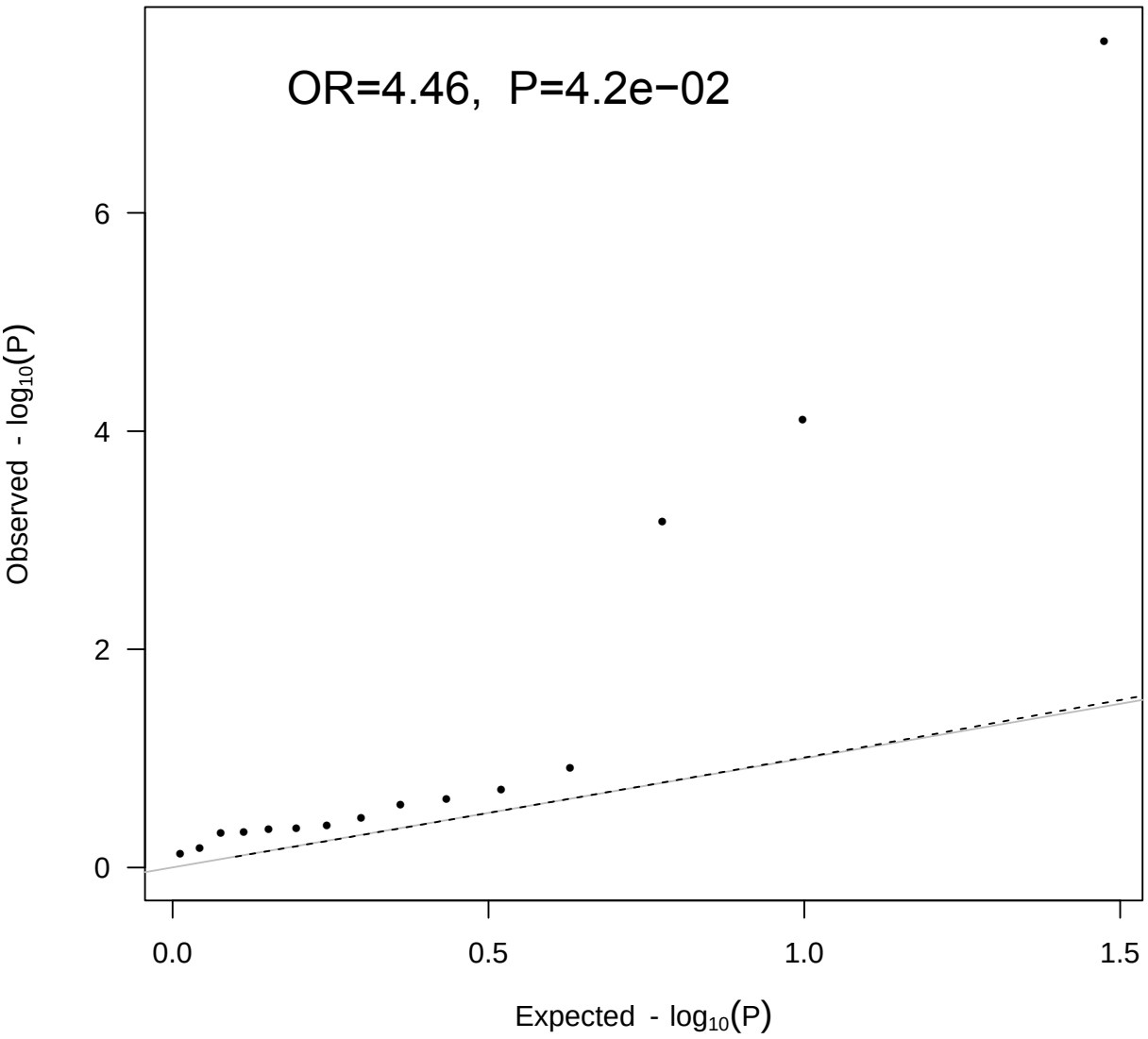
Ulcerative colitis:



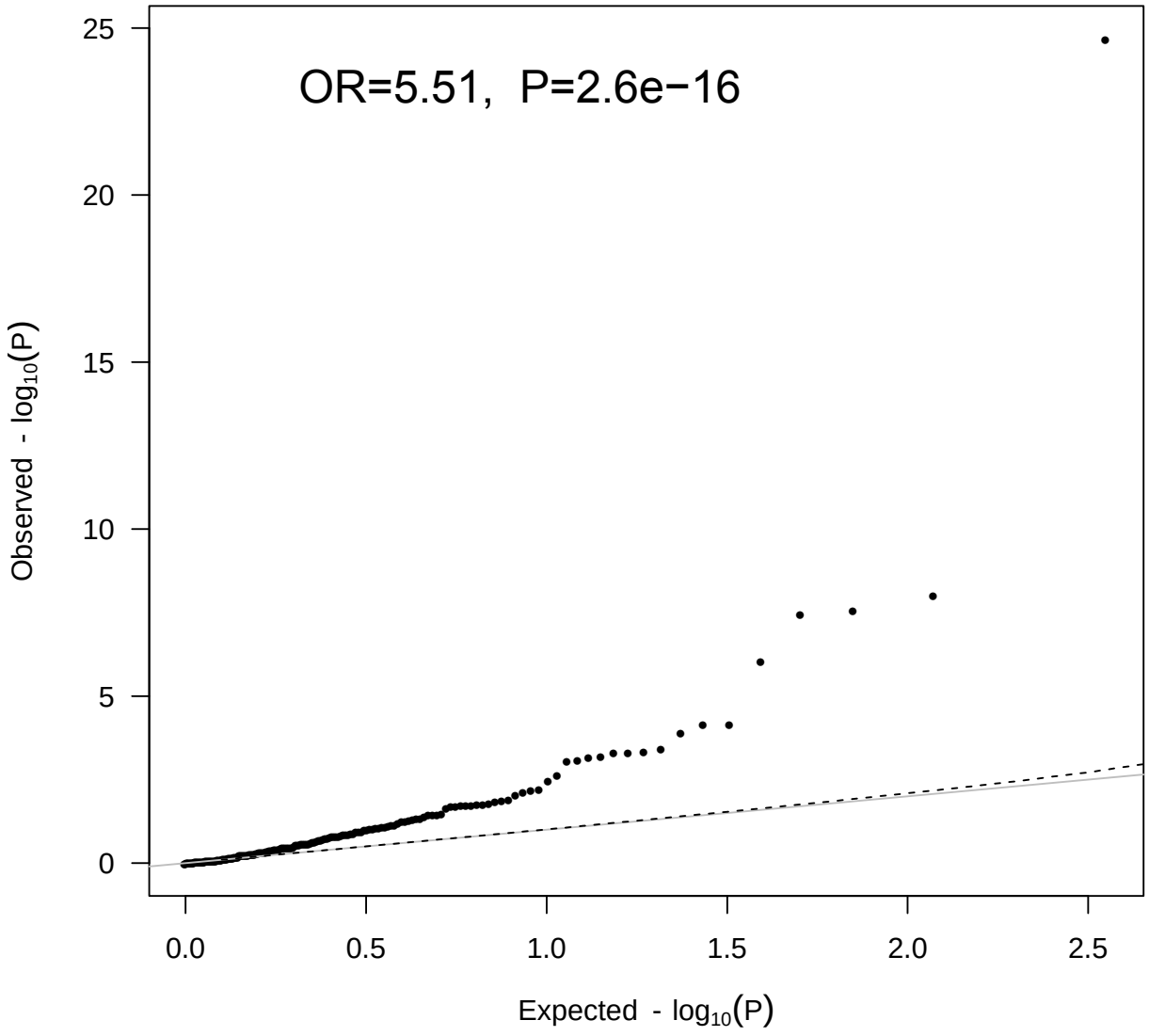
Crohn's Disease:



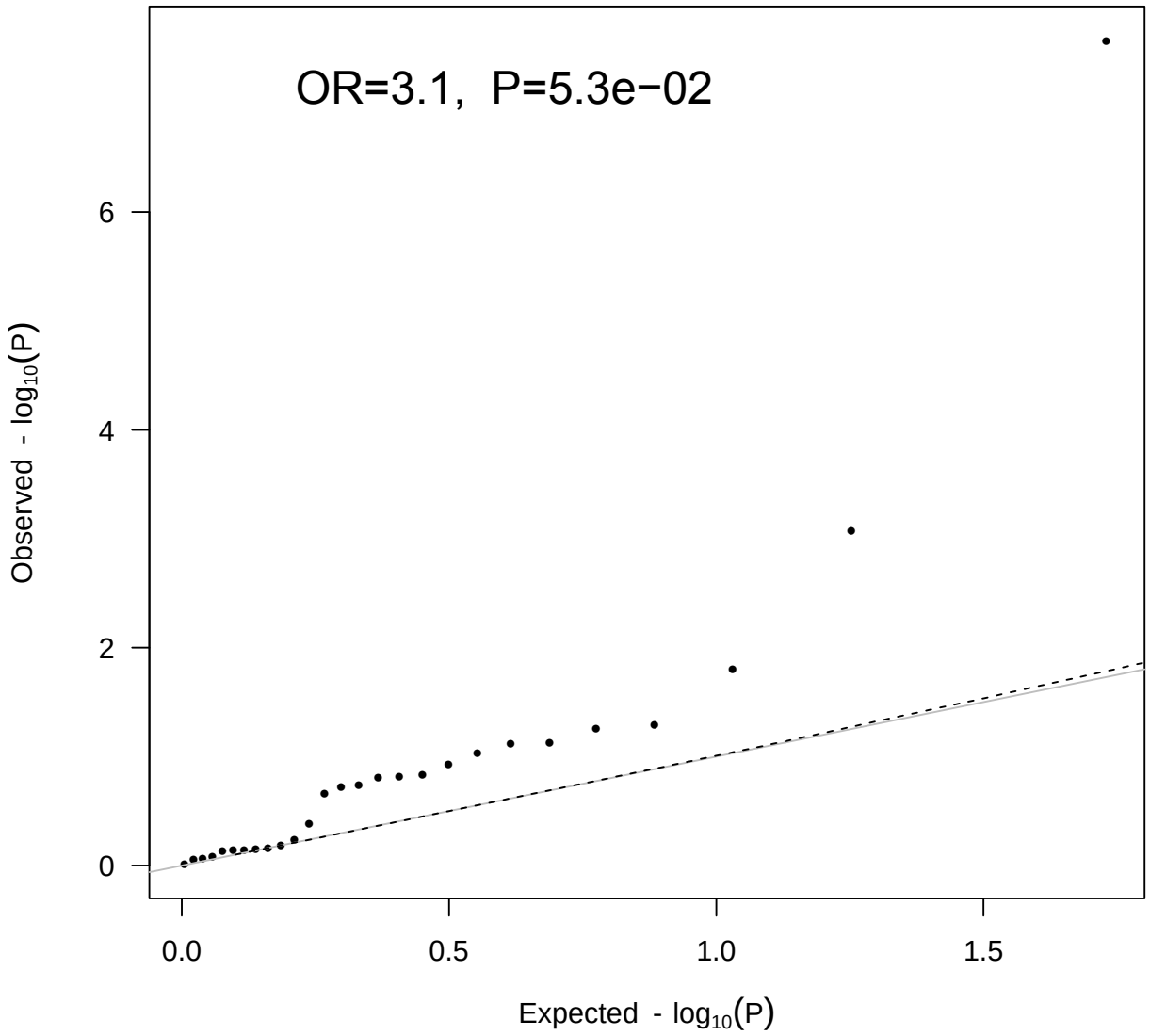
Graves Disease:



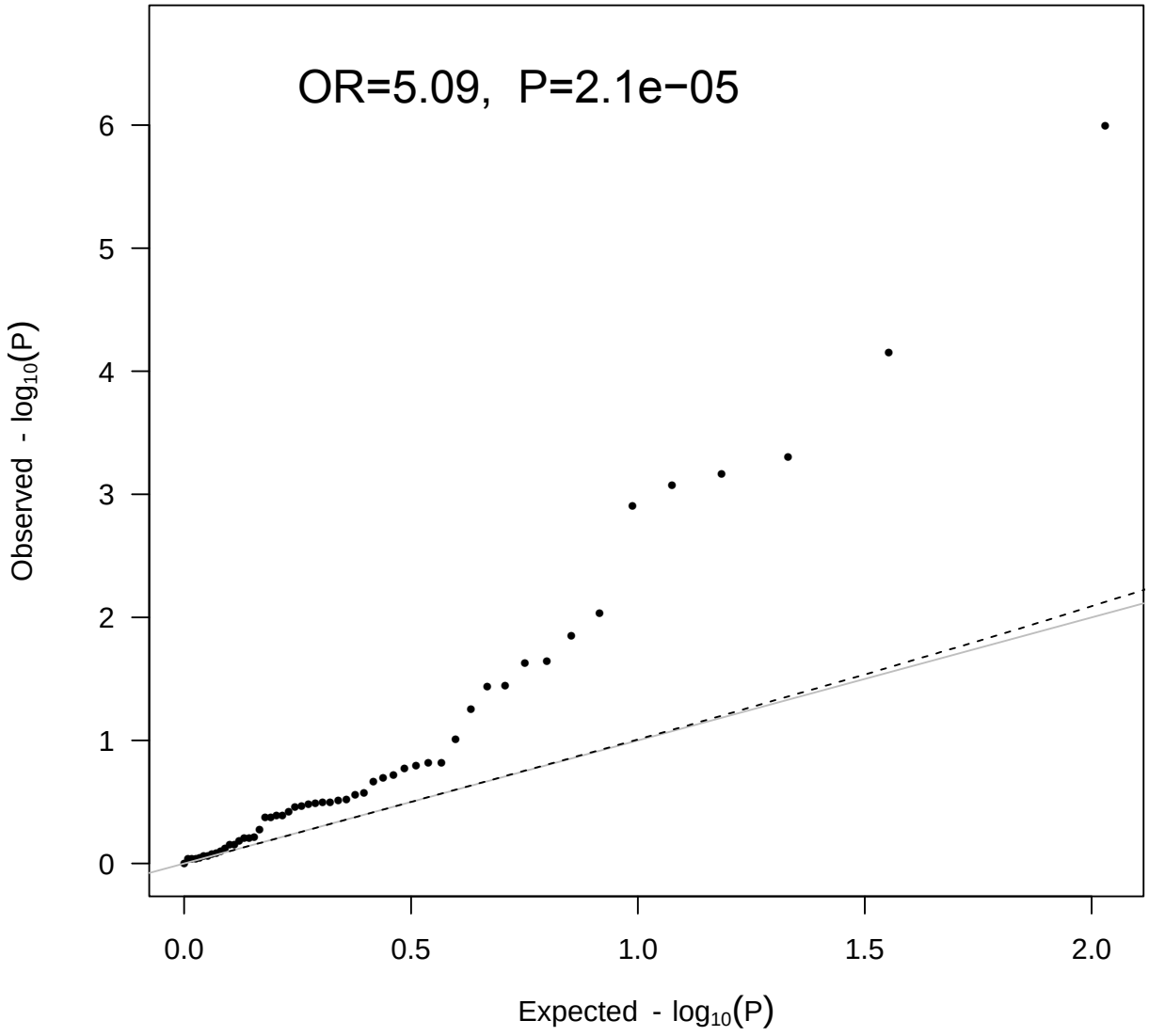
Inflammatory Bowel Disease:



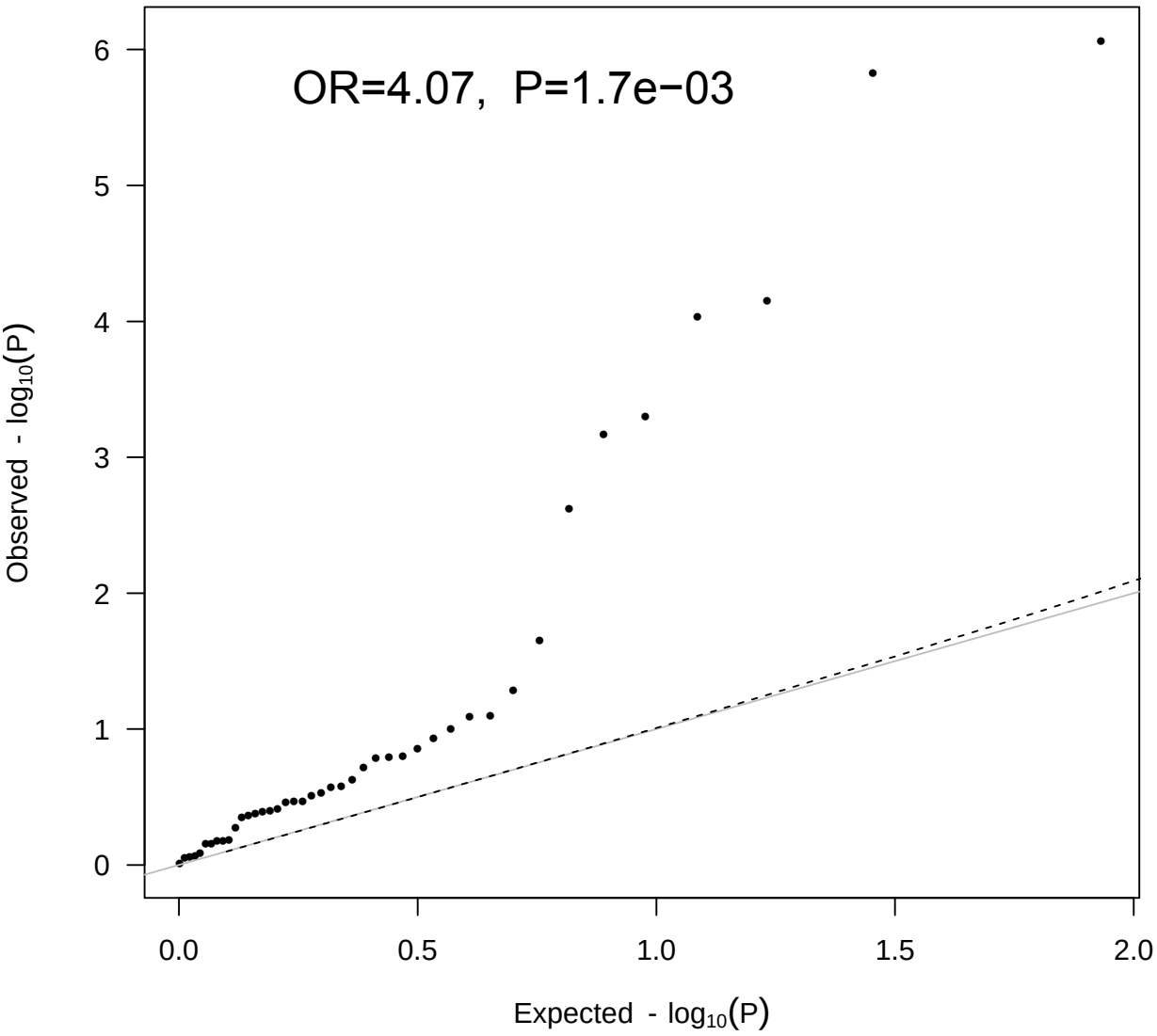
Systemic Lupus Erythematosus:



Multiple Sclerosis:



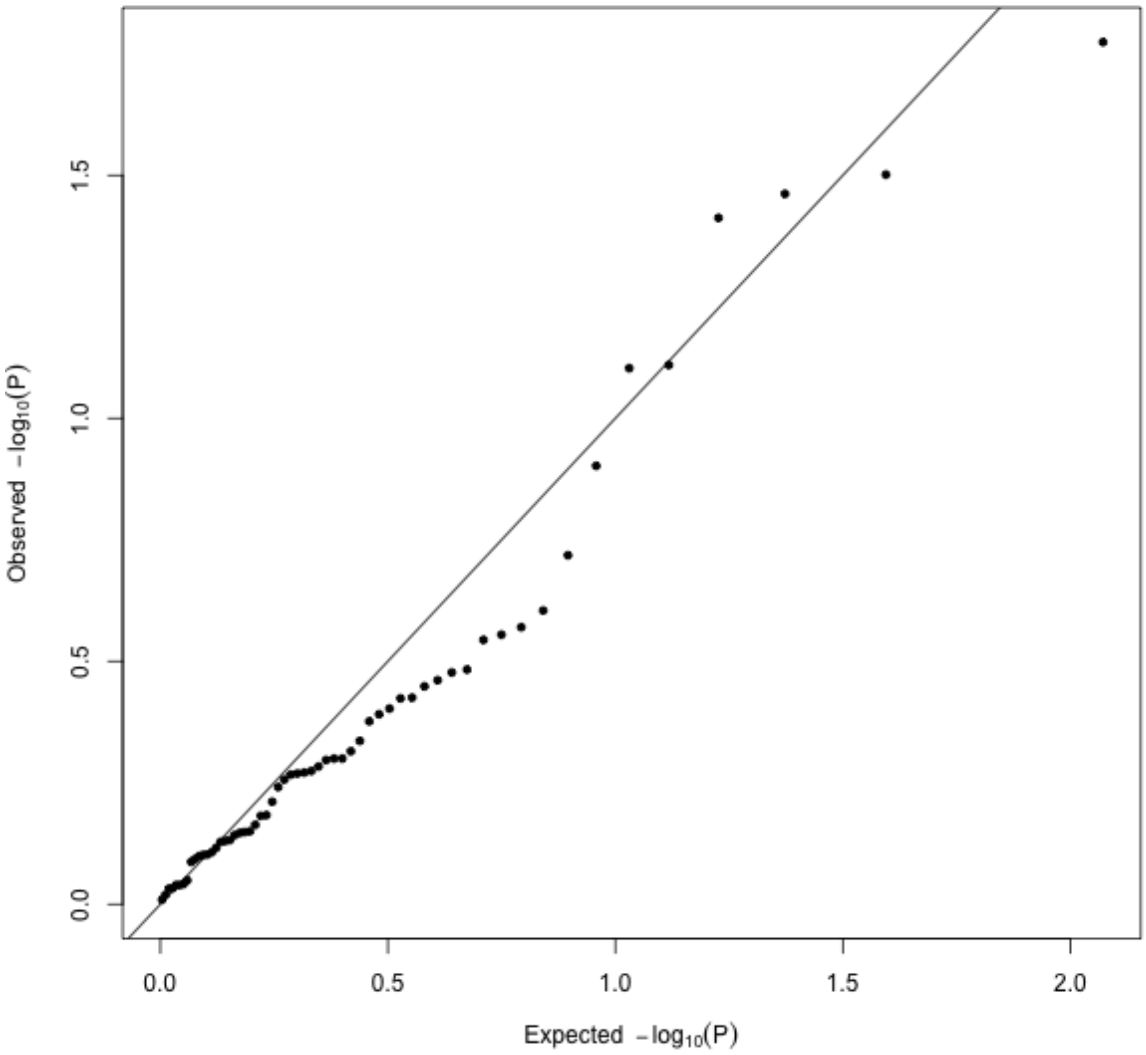
Type-1 Diabetes:



Supplementary Figure 7

QQ plot of 57 Migraine loci extracted from the allergy meta-analysis results.

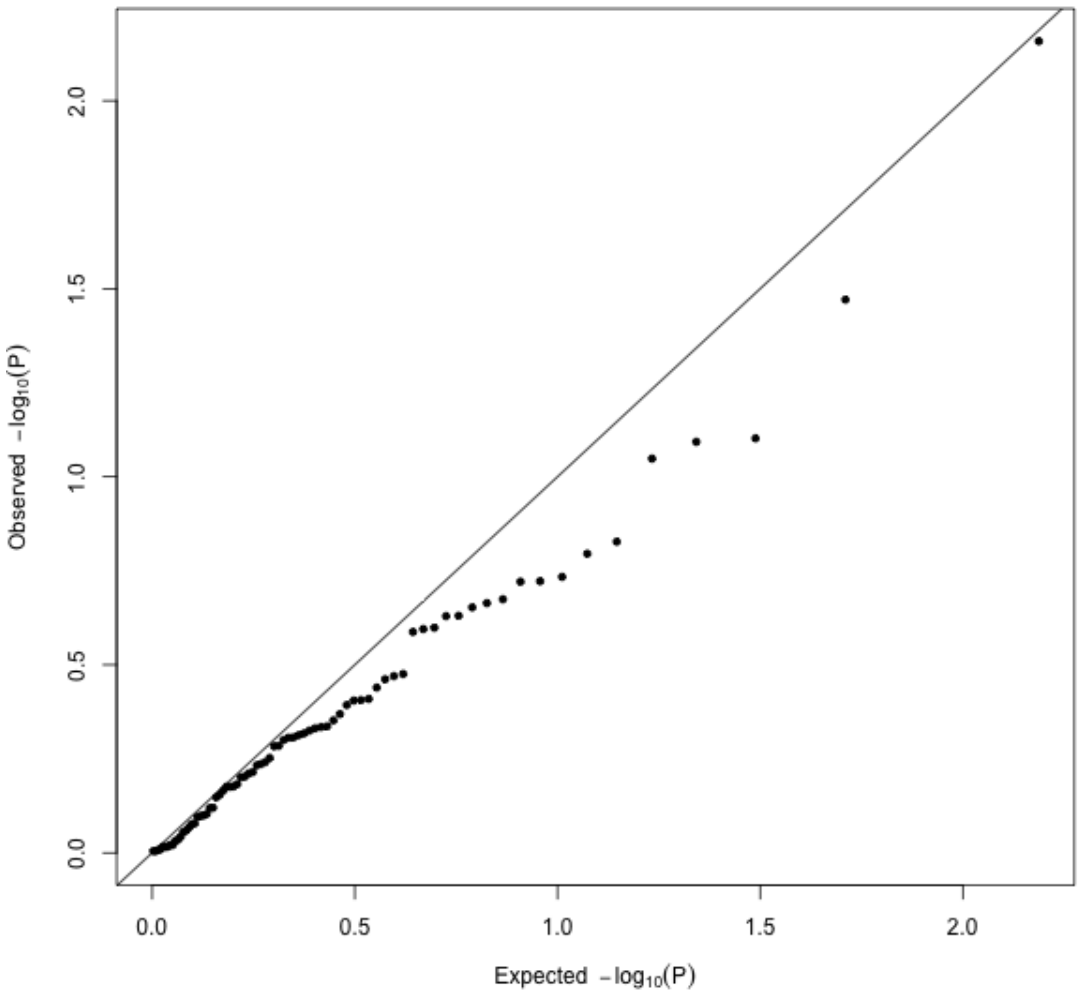
Migraine:



Supplementary Figure 8

QQ plot of 77 loci associated with the combined phenotype of schizophrenia and bipolar disorder extracted from the allergy meta-analysis results.

Bipolar disorder and schizophrenia:



Supplementary Figure 9

Principal component plot of GWAS Catalogue SNPs' perturbation of gene networks, based on the DEPICT tool, PC1 vs PC3

(PLEASE SEE SEPARATE FILE)

Supplementary Figure 10

Principal component plot of GWAS Catalogue SNPs' perturbation of gene networks, based on the DEPICT tool, PC1 vs PC2, all trait names.

(PLEASE SEE SEPARATE FILE)

Supplementary Figure 11

Allergy related loci and their resemblance to autoimmune disease and other types of disease loci were assessed by principal component analysis by analyzing the tendency of each trait-locus to fall in DHS sites in specific cell lines. This plot shows PC1 vs. PC2 and has the outlier “lipid metabolism phenotypes” omitted, and only names for autoimmune diseases, asthma and allergy are printed. The blue area represents the shared minimal ellipsoid area of immune-mediated diseases.

(PLEASE SEE SEPARATE FILE)

Supplementary Figure 12

Allergy related loci and their resemblance to autoimmune disease and other types of disease loci were assessed by principal component analysis by analyzing the tendency of each trait-locus to fall in DHS sites in specific cell lines. This plot shows PC1 vs. PC2 for the full data set.

(PLEASE SEE SEPARATE FILE)

Supplementary Figure 13

Allergy related loci and their resemblance to autoimmune disease and other types of disease loci were assessed by principal component analysis by analyzing the tendency of each trait-locus to fall in DHS sites in specific cell lines. This plot shows PC1 vs. PC2 overlayed with cell- and tissue type loadings.

(PLEASE SEE SEPARATE FILE)

Supplementary Figure 14

Hierarchical clustering of all NHRGI GWAS catalog diseases' associated SNPs' tendency to fall within DHS sites for immune cell types within the Encode data set.

(PLEASE SEE SEPARATE FILE)

Supplementary Figure 15

Enrichment of DHS sites in SNPs associated to allergy and Crohn's disease.
X-axis denominates all SNPs associated to the given trait at –log10(p) <= x, and y gives the enrichment of DHS sites for a given cell/tissue-type for those SNPs, as compared to all SNPs (x=0). Immune cells are indicated in blue.

(PLEASE SEE SEPARATE FILE)

Supplementary Figure 16

Enrichment of SNPs falling in FANTOM enhancers

(PLEASE SEE SEPARATE FILE)

Supplementary Figure 17

PCA plot of DEPICT pathway perturbation analysis, showing names for all gene sets.

(PLEASE SEE SEPARATE FILE)

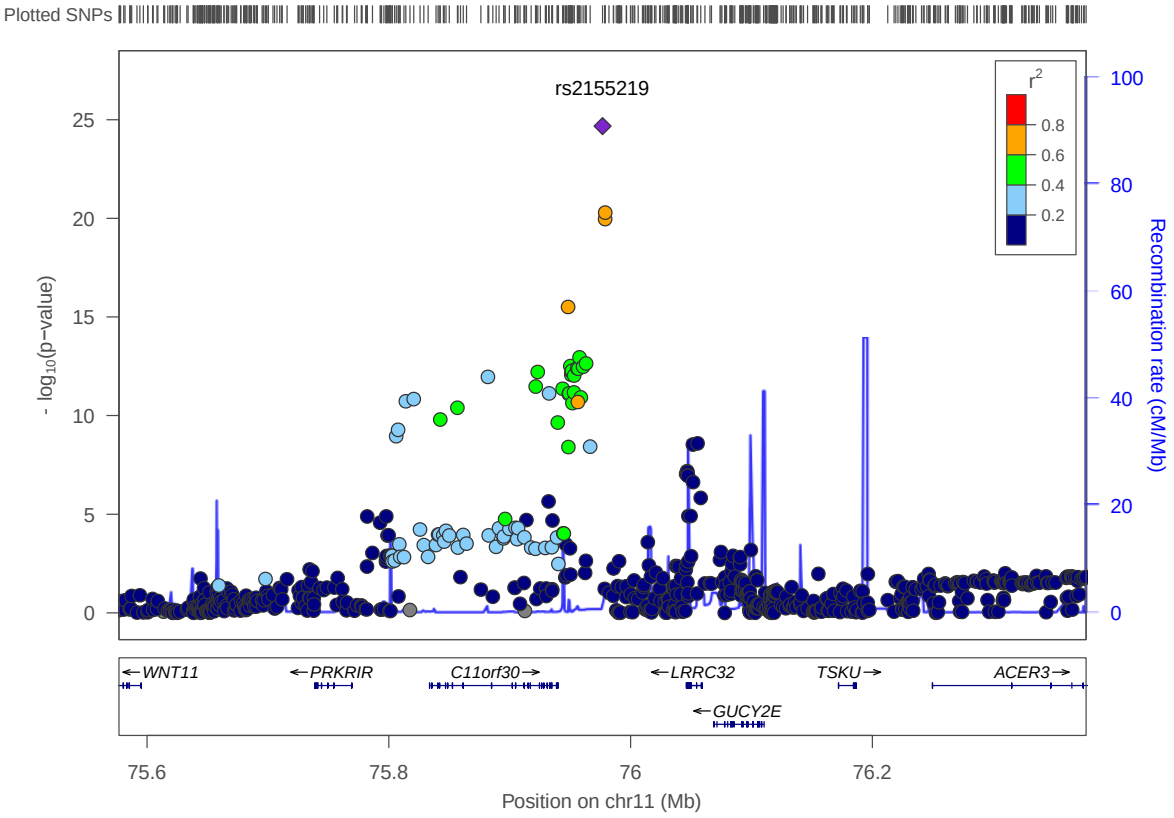
Supplementary Figure 18

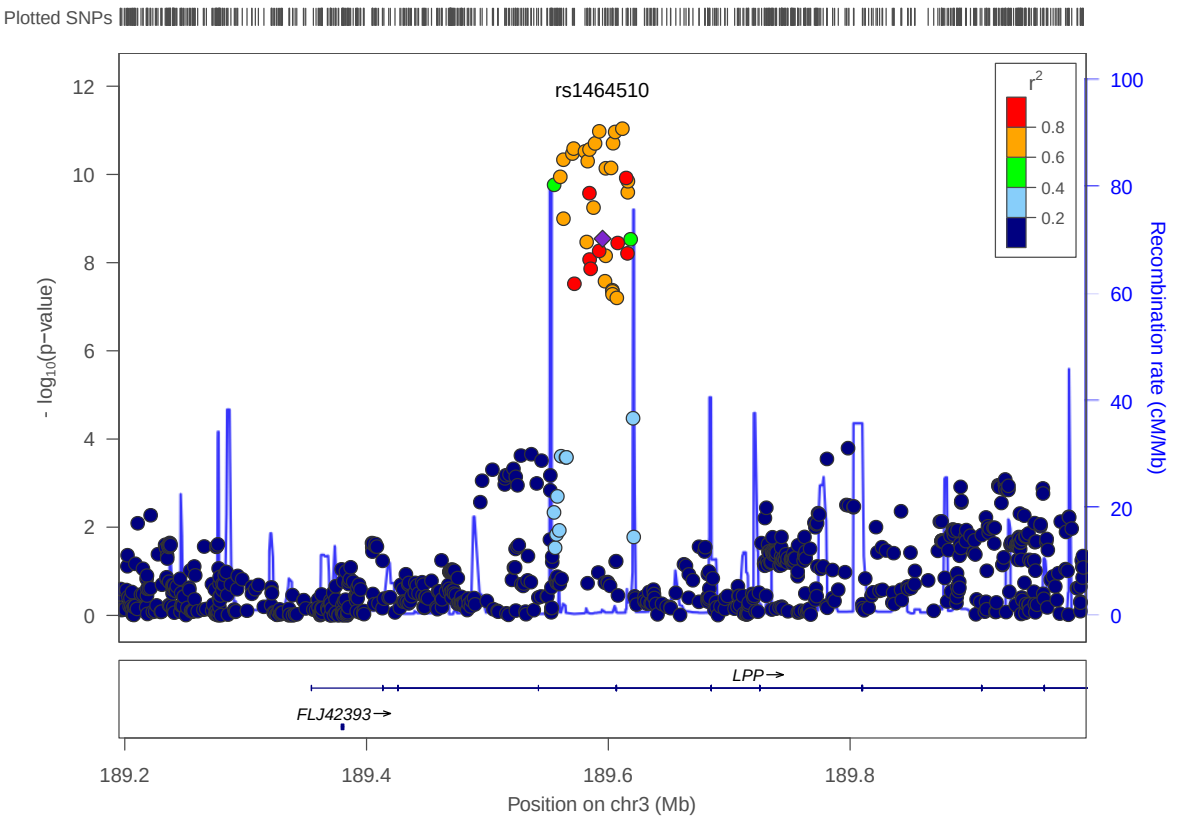
Enrichment of shared loci with ENCODE ChIP-seq based transcription factor binding sites. Green line indicates FDR < 0.05. Transcription factors in blue have FDR < 0.05 and enrichment >= 3.

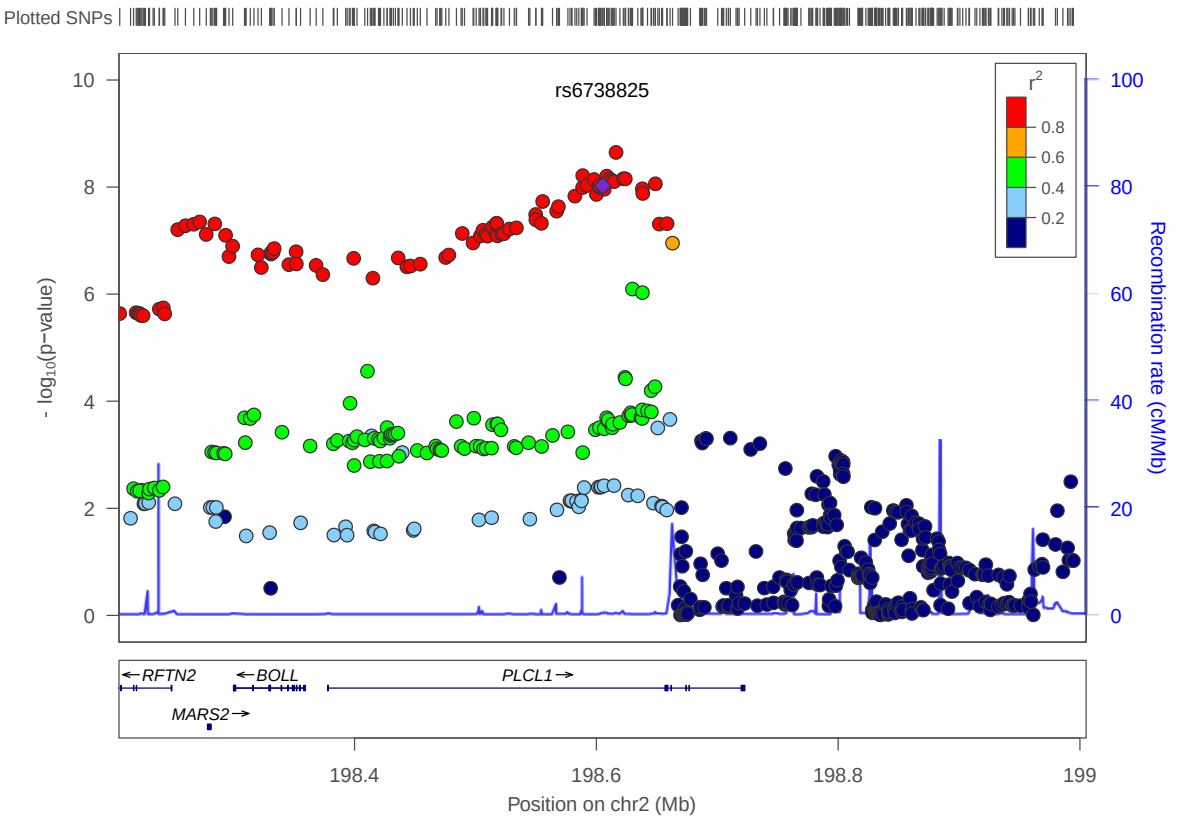
(PLEASE SEE SEPARATE FILE)

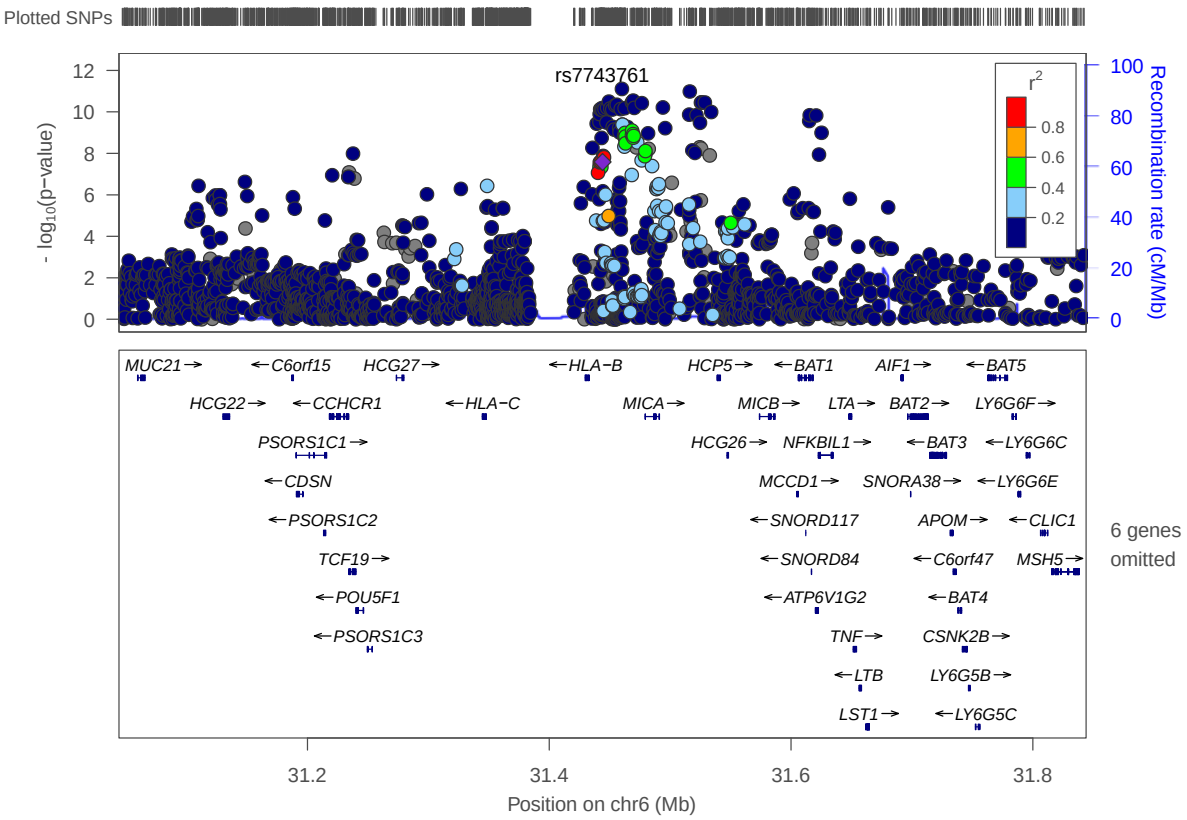
Supplementary Figure 19

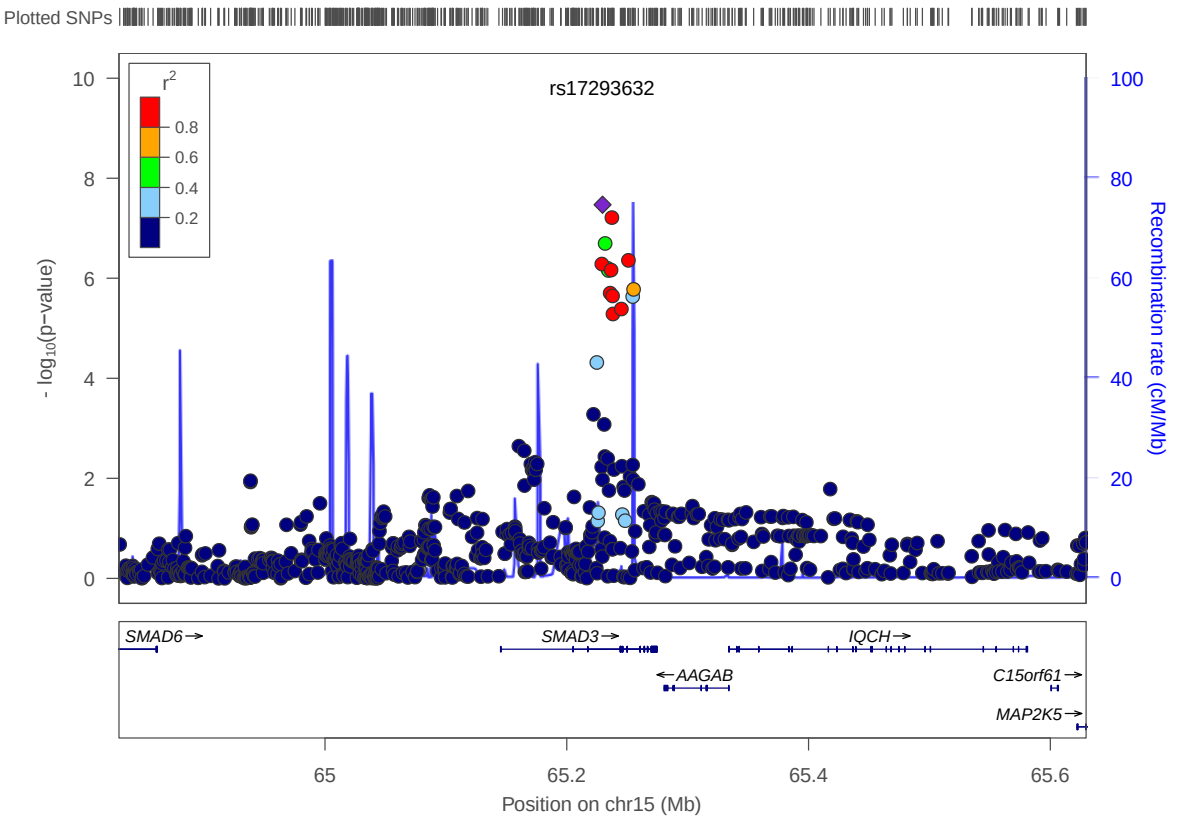
LocusZoom plots of the autoimmune disease associated loci within the allergy meta-analysis. Each dot represents the association between allergy and the particular SNP. The purple SNP is the index SNP for which the remaining SNPs are colored with respect to the r2 value to the index SNP. The position on the Y-axis represents the P-value (left handside Y-axis). The blue line represents recombination rates (righ handside Y-axis)

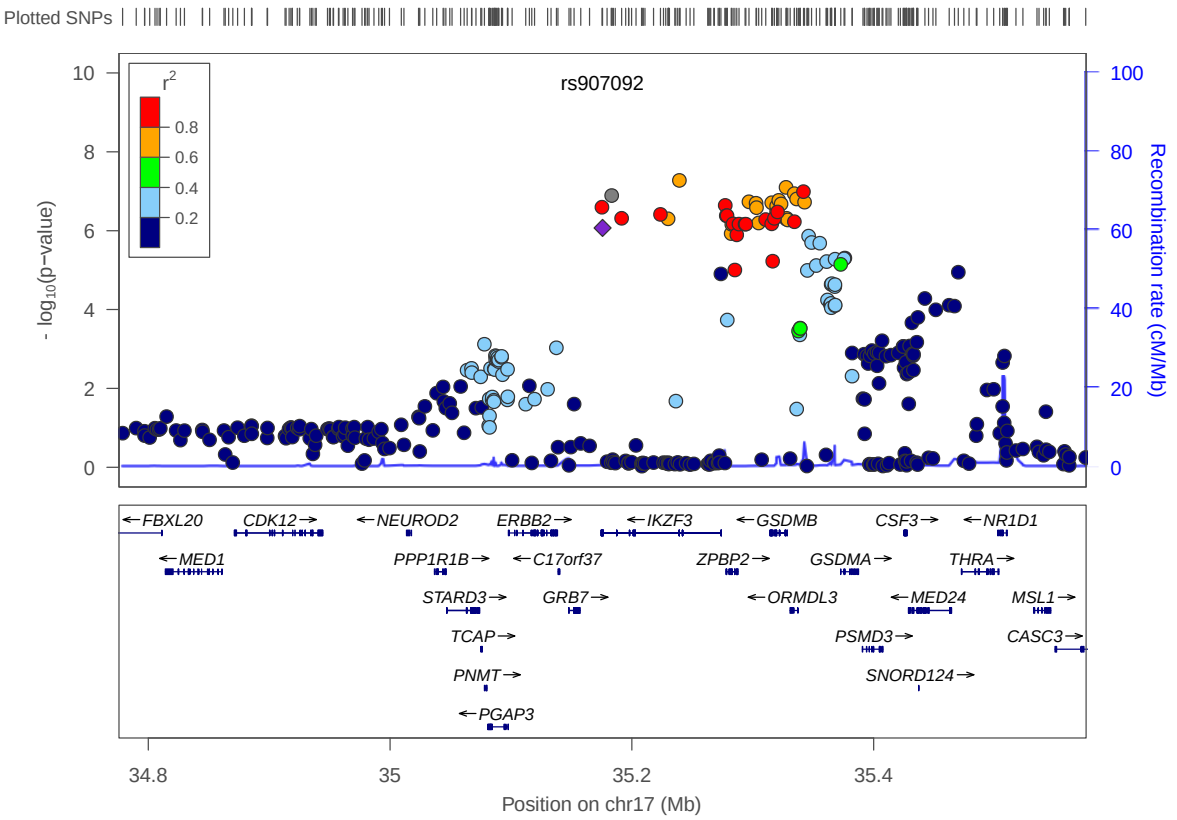


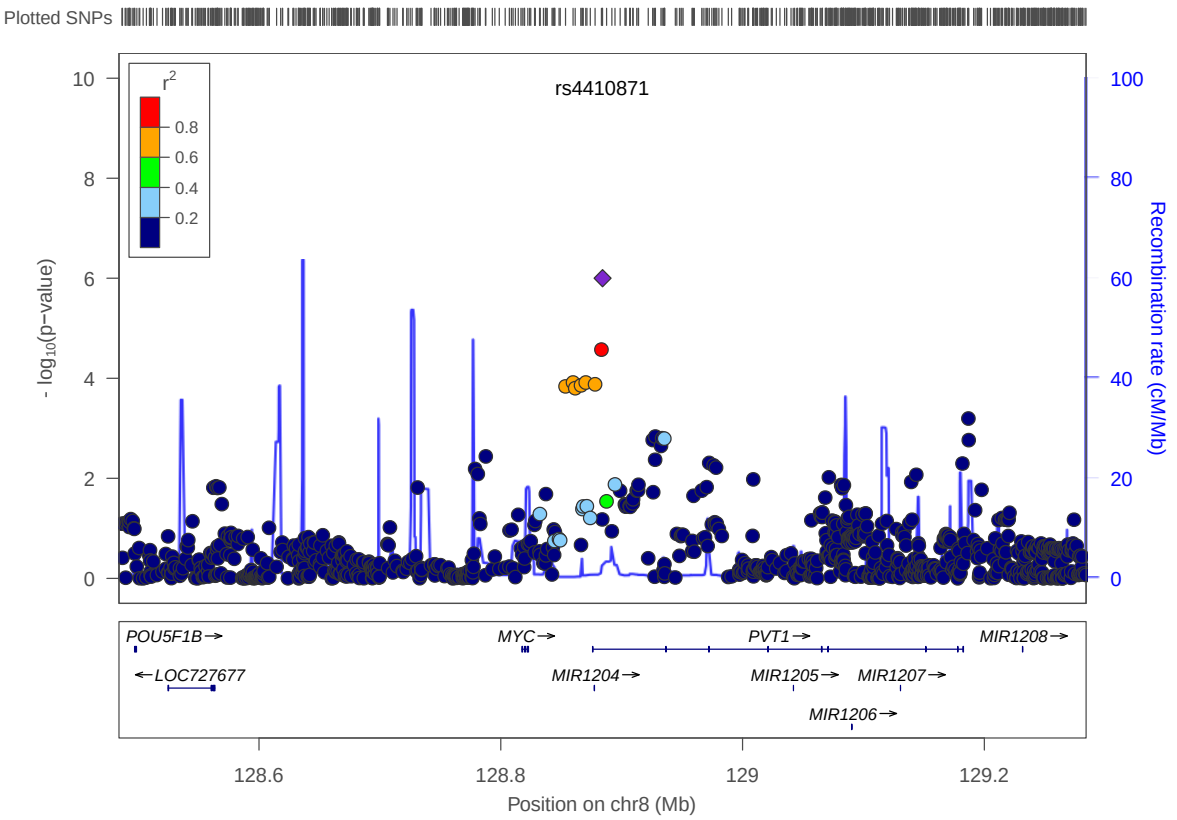

Plotted SNPs
rs1464510
r2
0.8
0.6
0.4
0.2
- log10(p-value)
Recombination rate (cM/Mb)
LPP
FLJ42393
189.2
189.4
189.6
189.8
Position on chr3 (Mb)

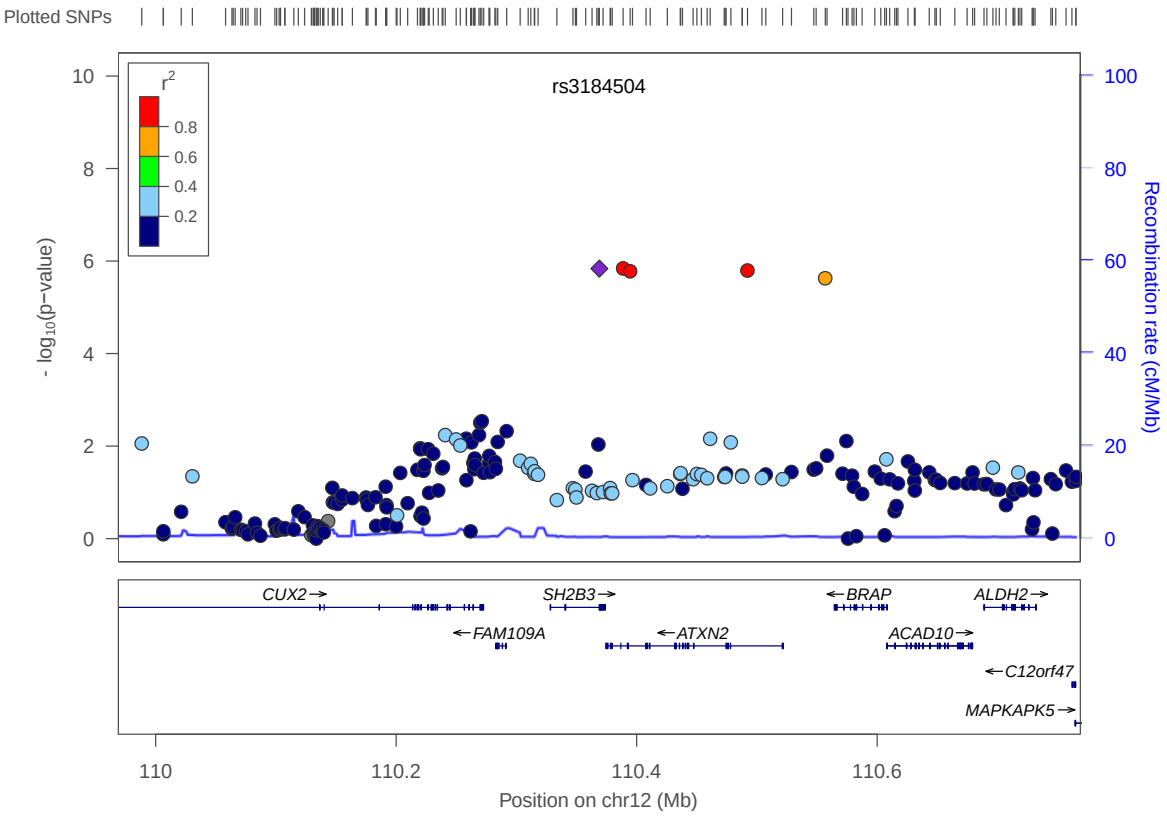

Plotted SNPs
rs6738825
r2
0.8
0.6
0.4
0.2
- log10(p-value)
Recombination rate (cM/Mb)
RFTN2
BOLL
PLCL1
MARS2
198.4
198.6
198.8
199
Position on chr2 (Mb)

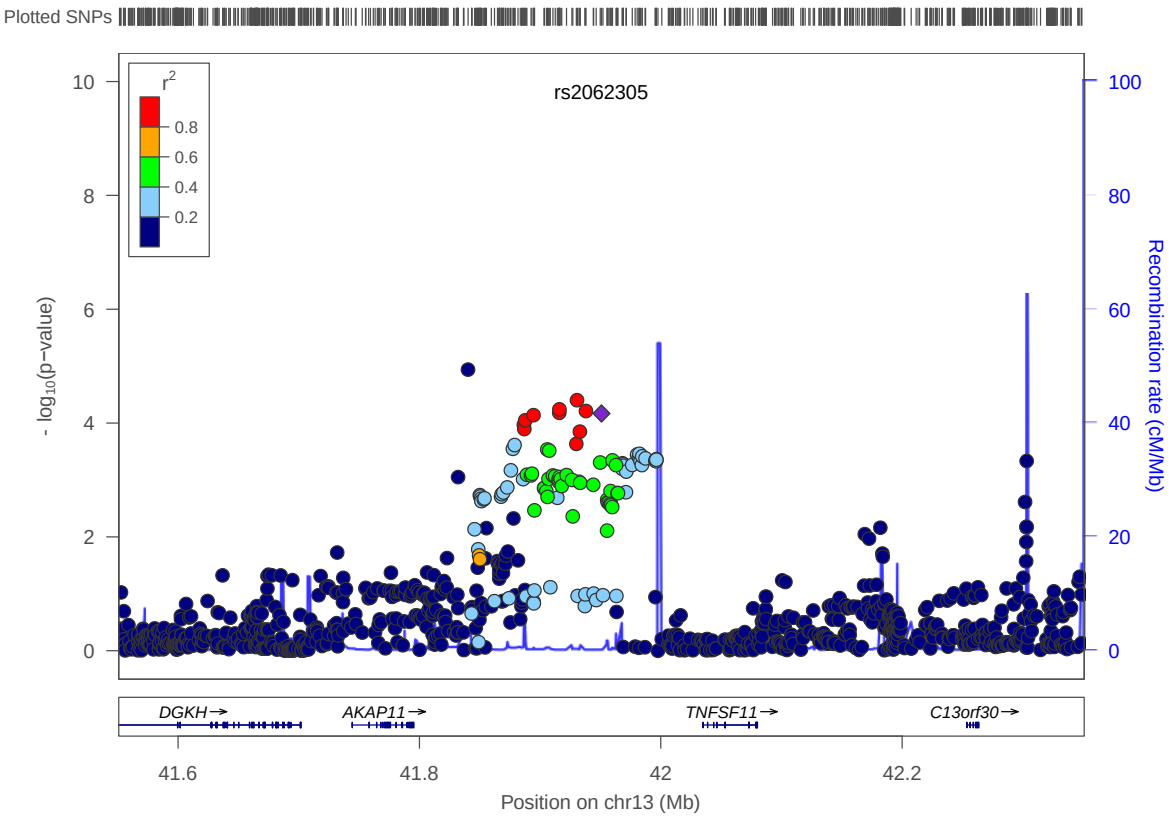

Plotted SNPs
rs7743761
r²
0.8
0.6
0.4
0.2
- log10(p-value)
Recombination rate (cM/Mb)
MUC21 C6orf15 HCG27 HLA-B HCP5 BAT1 AIF1 BAT5
HCG22 CCHCR1 HLA-C MICA MICB LTA BAT2 LY6G6F
PSORS1C1 HCG26 NFKBIL1 BAT3 LY6G6C
CDSN MCCD1 SNORA38 LY6G6E
PSORS1C2 SNORD117 APOM CLIC1
TCF19 SNORD84 C6orf47 MSH5
POU5F1 ATP6V1G2 BAT4
PSORS1C3 TNF CSNK2B
LTB LY6G5B
LST1 LY6G5C
6 genes omitted
31.2
31.4
31.6
31.8
Position on chr6 (Mb)

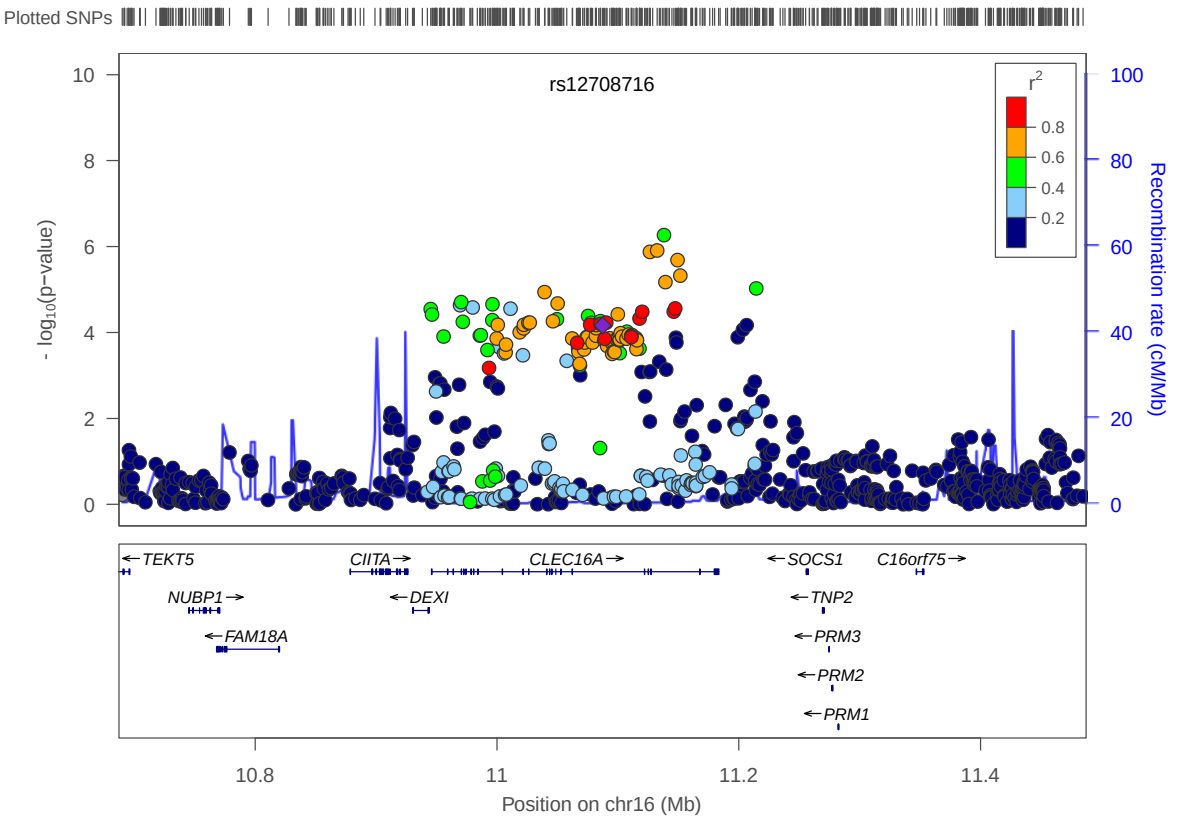

Plotted SNPs
rs17293632
r2
0.8
0.6
0.4
0.2
- log10(p−value)
Recombination rate (cM/Mb)
SMAD6
SMAD3
IQCH
AAGAB
C15orf61
MAP2K5
Position on chr15 (Mb)

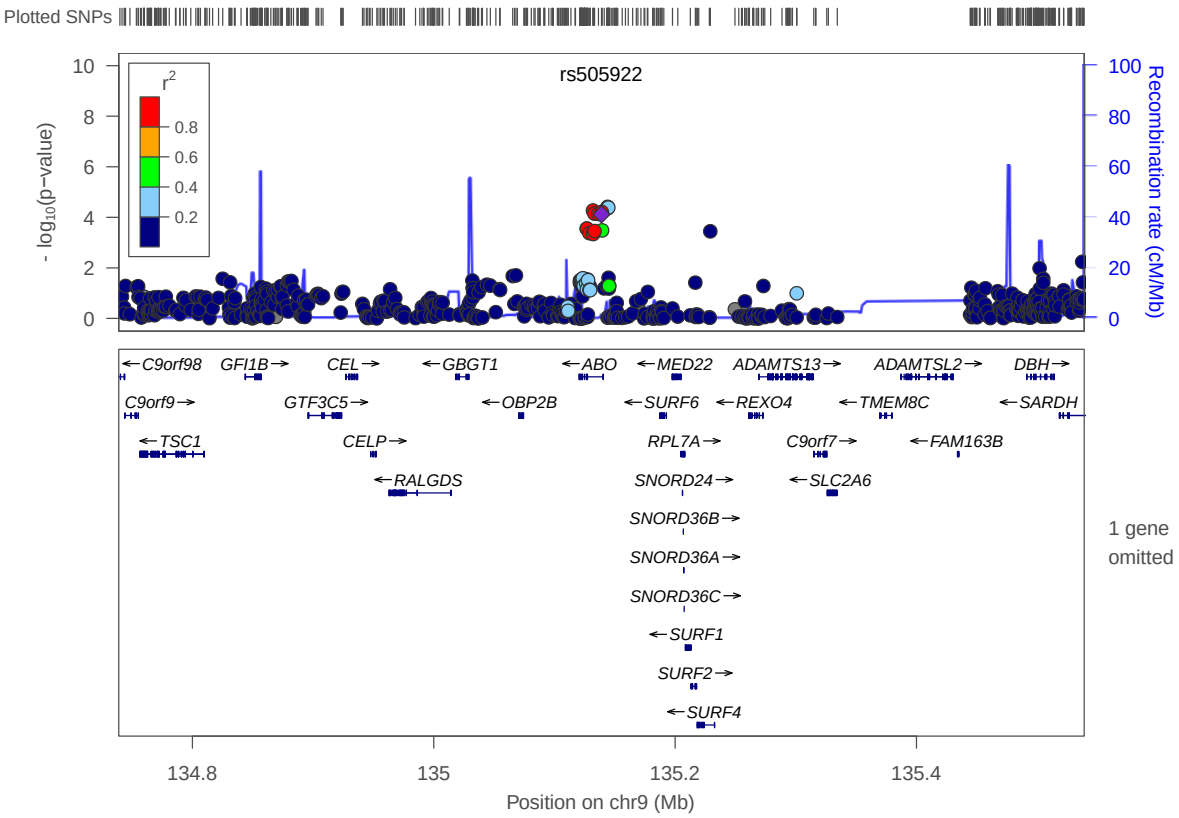

Plotted SNPs
rs907092
r2
0.8
0.6
0.4
0.2
- log10(p-value)
Recombination rate (cM/Mb)
FBXL20
CDK12
NEUROD2
ERBB2
IKZF3
GSDMB
CSF3
NR1D1
MED1
PPP1R1B
C17orf37
ZPBP2
GSDMA
THRA
STARD3
GRB7
ORMDL3
MED24
MSL1
TCAP
PSMD3
CASC3
PNMT
SNORD124
PGAP3
Position on chr17 (Mb)

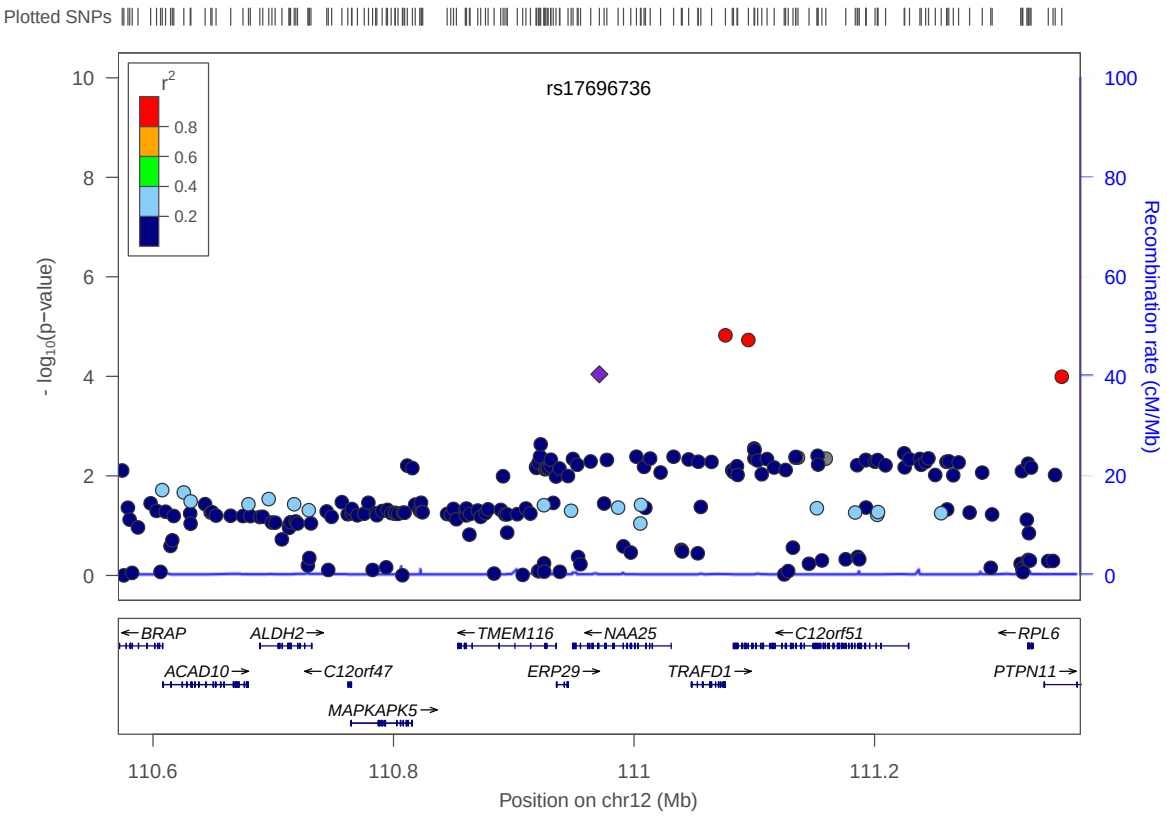

Plotted SNPs
rs4410871
r²
0.8
0.6
0.4
0.2
- log10(p−value)
Recombination rate (cM/Mb)
POU5F1B
MYC
PVT1
MIR1208
LOC727677
MIR1204
MIR1205
MIR1207
MIR1206
128.6
128.8
129
129.2
Position on chr8 (Mb)

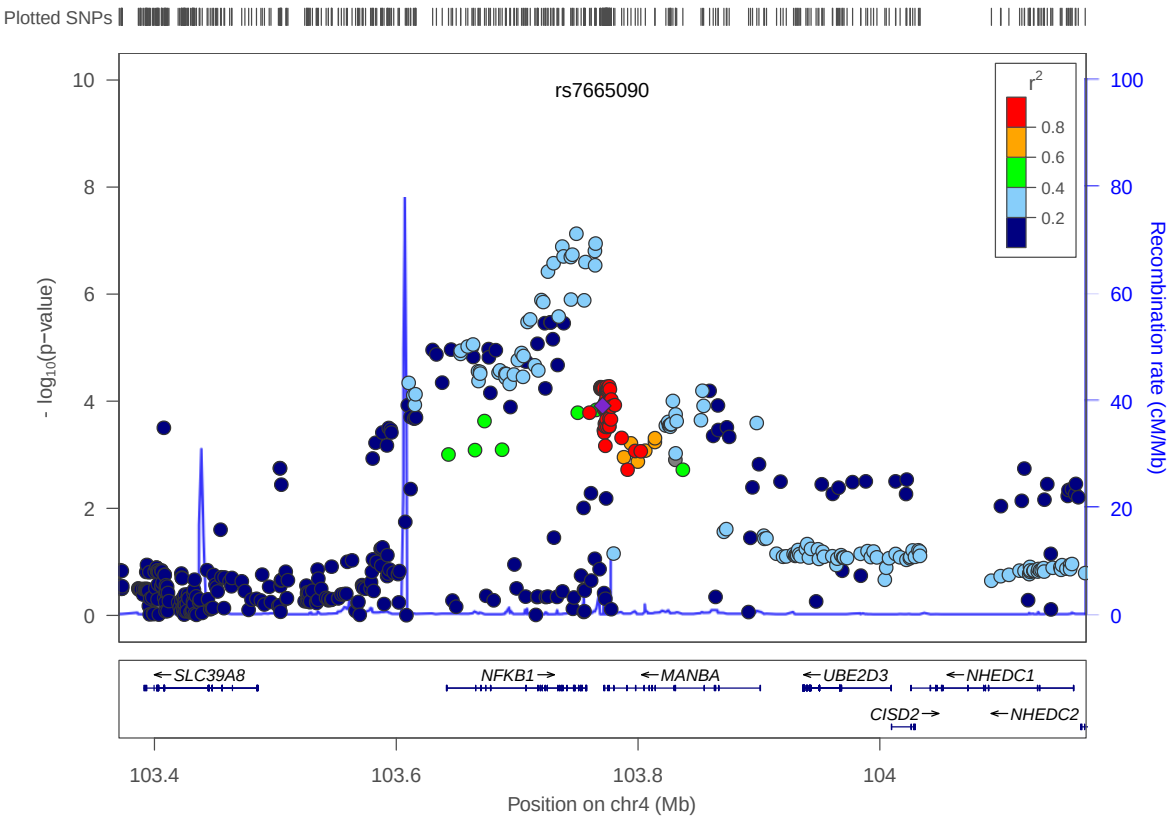

Plotted SNPs
rs3184504
r2
0.8
0.6
0.4
0.2
- log10(p-value)
Recombination rate (cM/Mb)
CUX2
SH2B3
BRAP
ALDH2
FAM109A
ATXN2
ACAD10
C12orf47
MAPKAPK5
Position on chr12 (Mb)

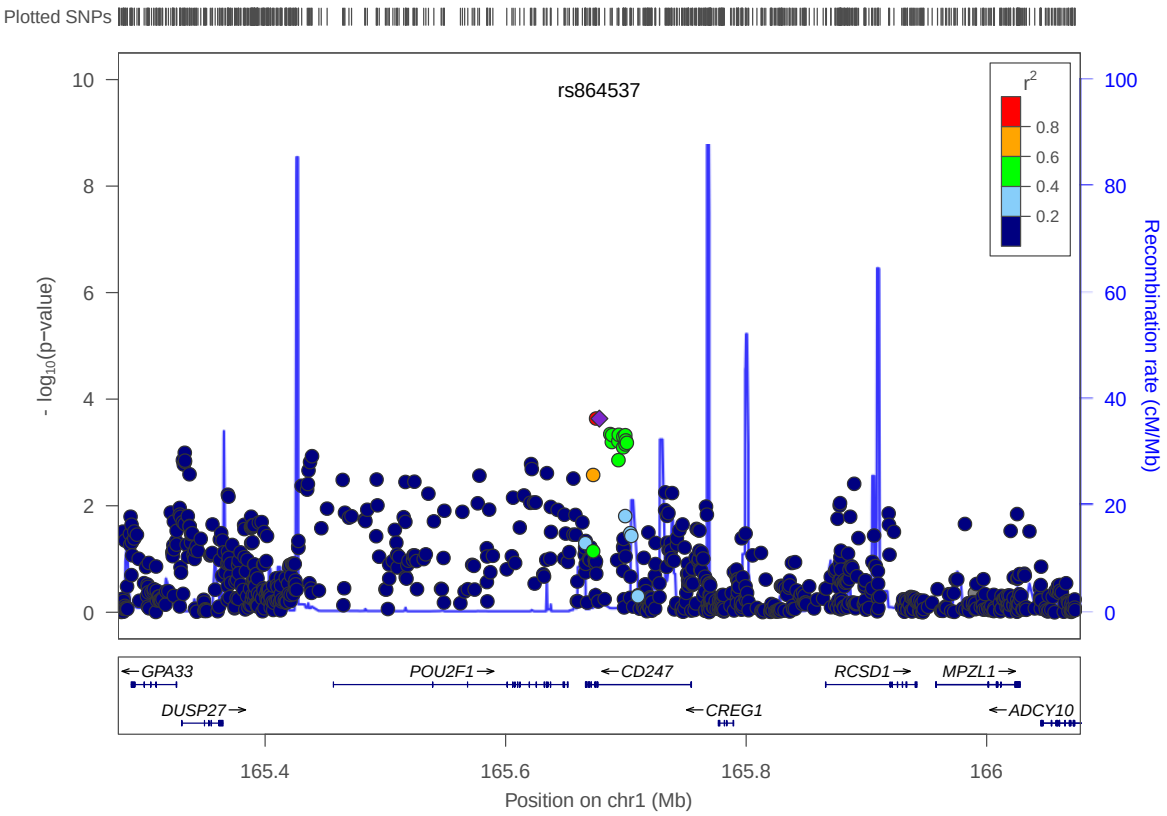

Plotted SNPs
rs2062305
r2
0.8
0.6
0.4
0.2
- log10(p-value)
Recombination rate (cM/Mb)
DGKH
AKAP11
TNFSF11
C13orf30
Position on chr13 (Mb)

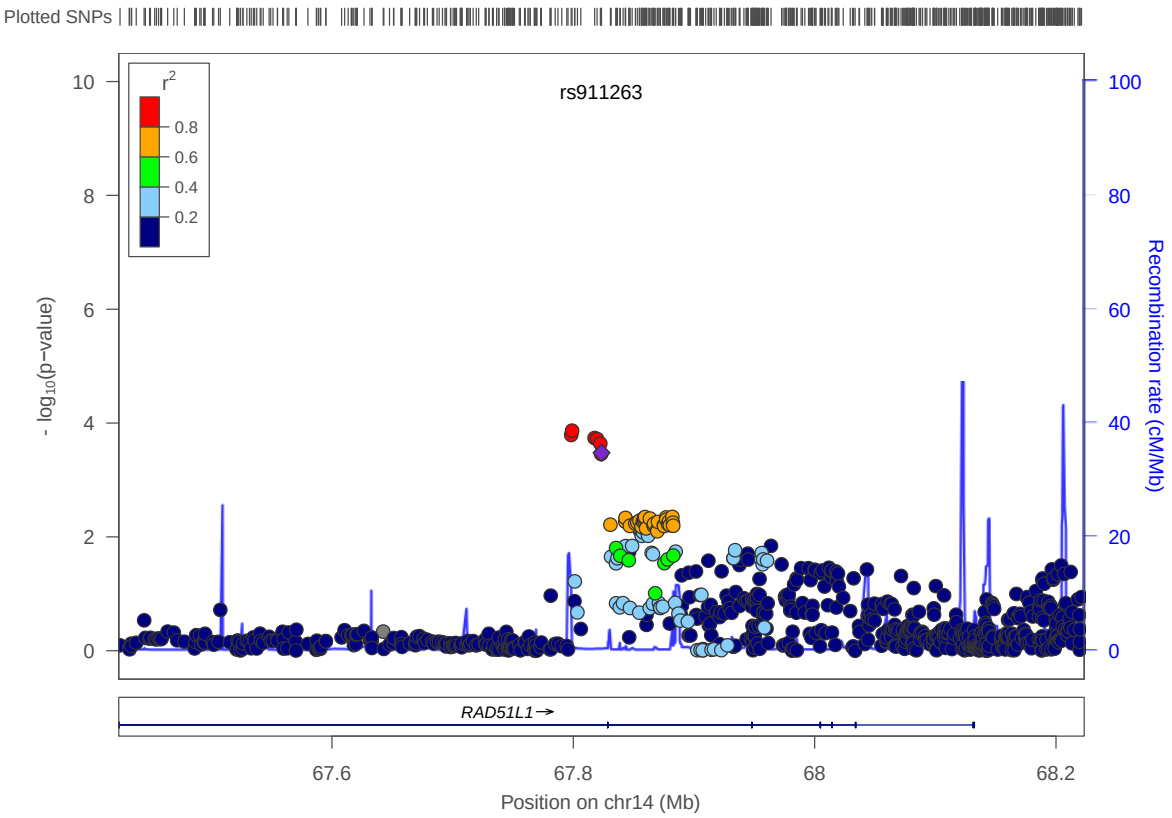

Plotted SNPs
rs12708716
r²
0.8
0.6
0.4
0.2
- log10(p-value)
Recombination rate (cM/Mb)
TEKT5
CIITA
CLEC16A
SOCS1
C16orf75
NUBP1
DEXI
TNP2
FAM18A
PRM3
PRM2
PRM1
Position on chr16 (Mb)

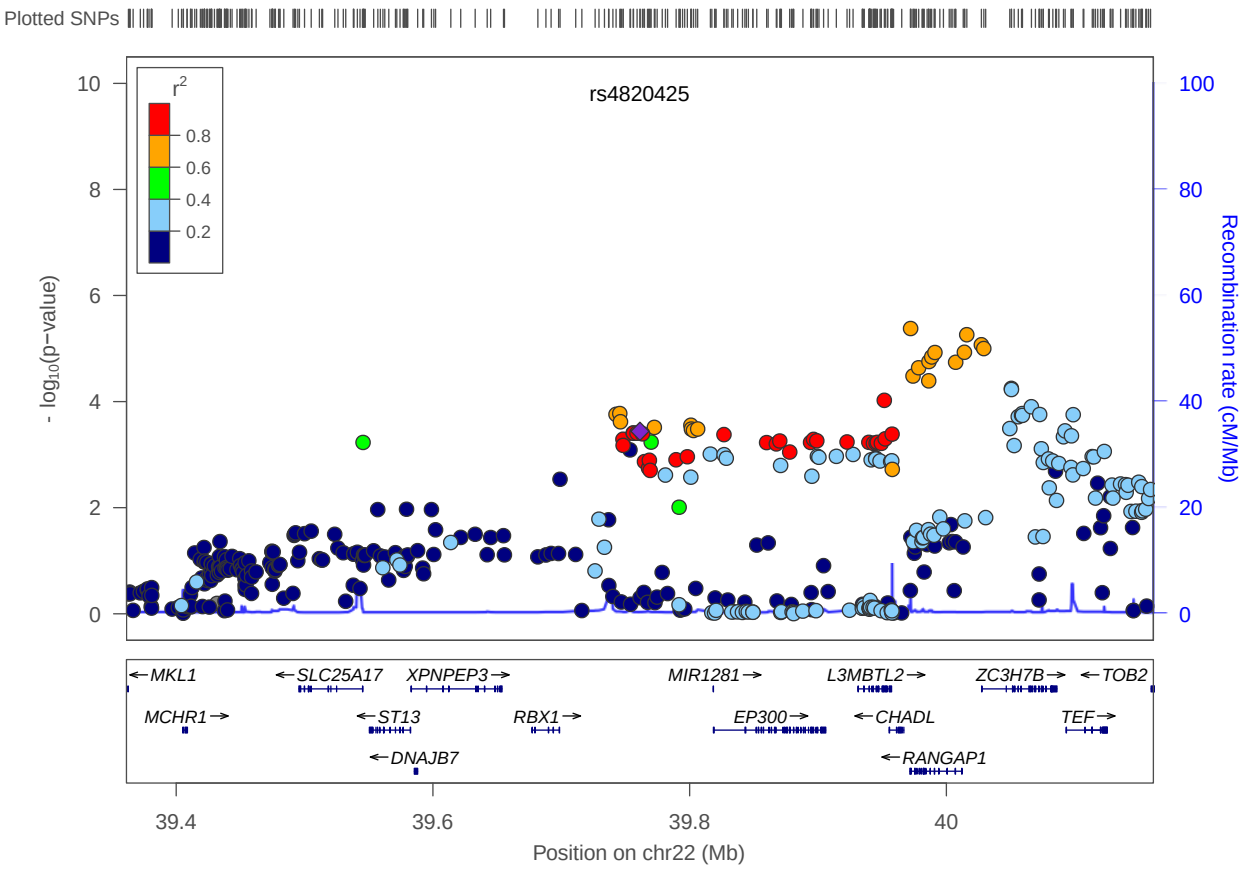

Plotted SNPs
rs505922
r²
0.8
0.6
0.4
0.2
- log10(p-value)
Recombination rate (cM/Mb)
C9orf98
GFI1B
CEL
GBGT1
ABO
MED22
ADAMTS13
ADAMTSL2
DBH
C9orf9
GTF3C5
OBP2B
SURF6
REXO4
TMEM8C
SARDH
TSC1
CELP
RPL7A
C9orf7
FAM163B
RALGDS
SNORD24
SLC2A6
SNORD36B
SNORD36A
SNORD36C
SURF1
SURF2
SURF4
1 gene omitted
134.8
135
135.2
135.4
Position on chr9 (Mb)

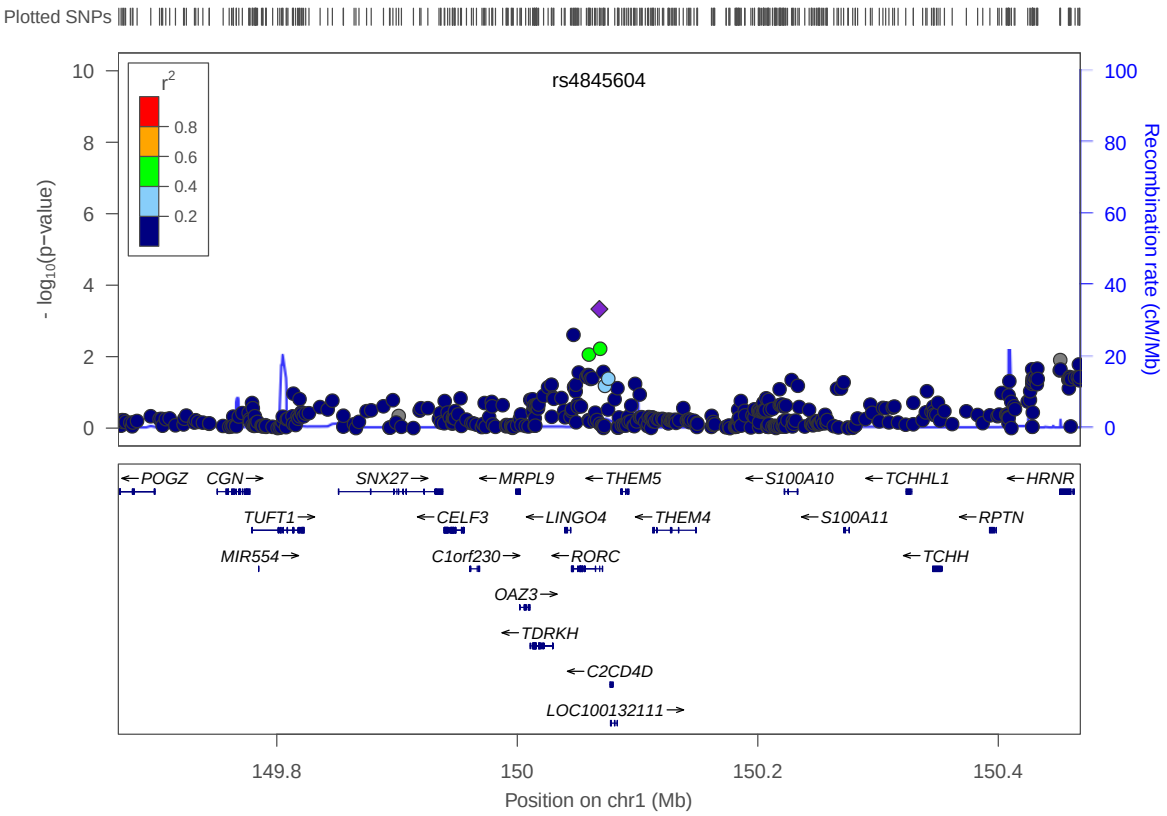

Plotted SNPs
rs17696736
r2
0.8
0.6
0.4
0.2
- log10(p−value)
Recombination rate (cM/Mb)
BRAP
ALDH2
TMEM116
NAA25
C12orf51
RPL6
ACAD10
C12orf47
ERP29
TRAFD1
PTPN11
MAPKAPK5
110.6
110.8
111
111.2
Position on chr12 (Mb)

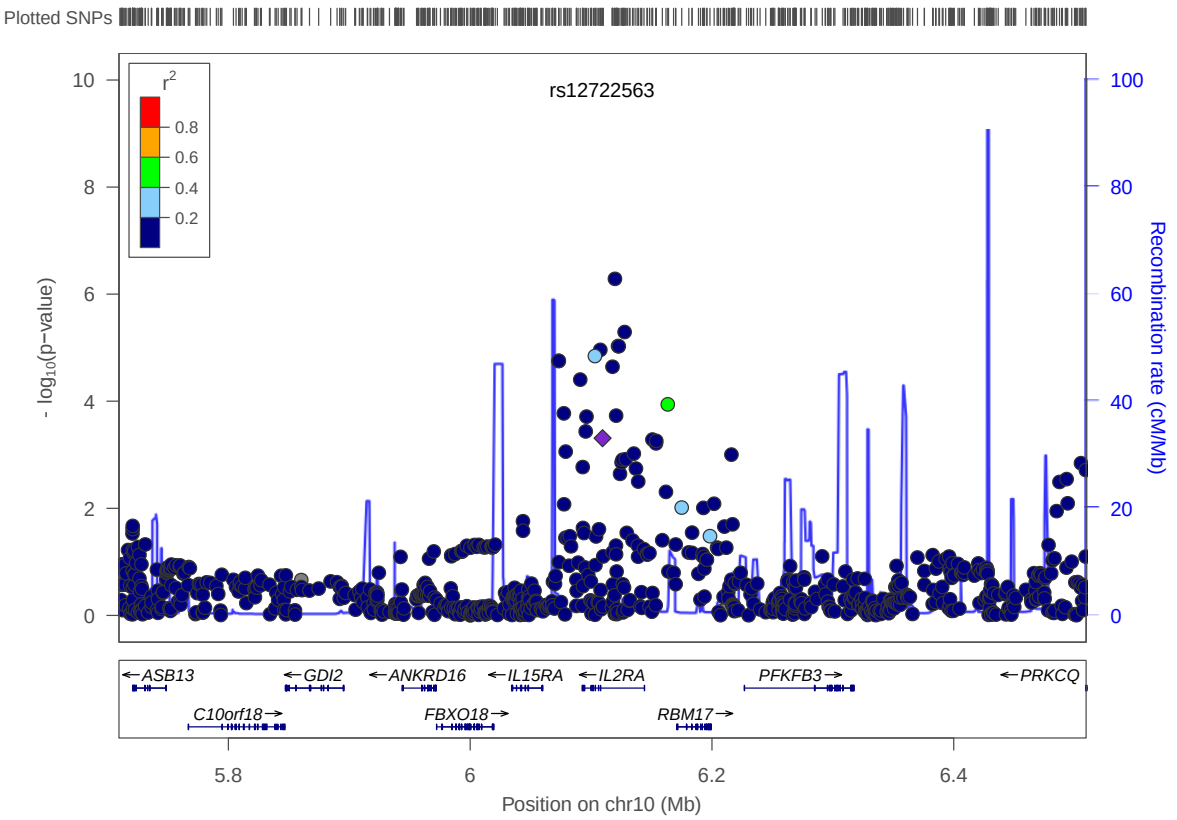

Plotted SNPs
rs7665090
r2
0.8
0.6
0.4
0.2
- log10(p-value)
Recombination rate (cM/Mb)
SLC39A8
NFKB1
MANBA
UBE2D3
NHEDC1
CISD2
NHEDC2
Position on chr4 (Mb)

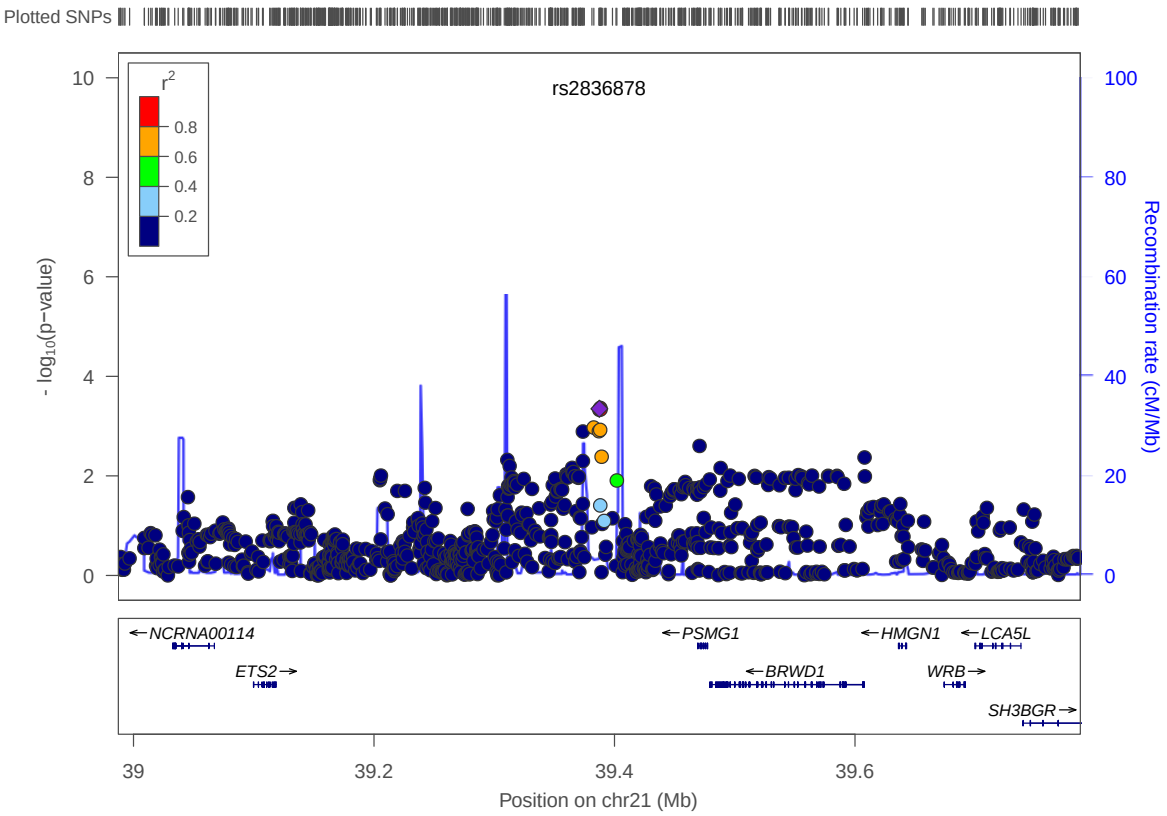

Plotted SNPs
rs864537
r2
0.8
0.6
0.4
0.2
- log10(p-value)
Recombination rate (cM/Mb)
GPA33
POU2F1
CD247
RCSD1
MPZL1
DUSP27
CREG1
ADCY10
165.4
165.6
165.8
166
Position on chr1 (Mb)

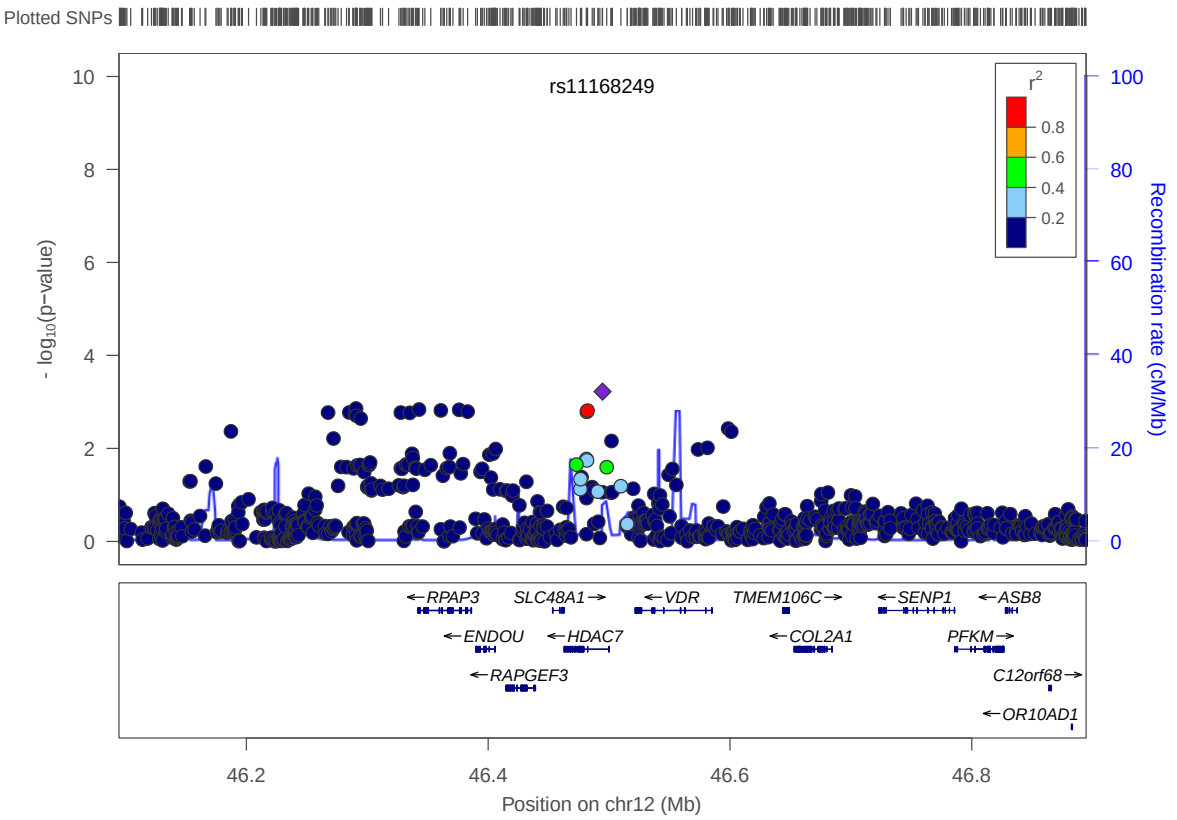

Plotted SNPs
rs911263
r2
0.8
0.6
0.4
0.2
- log10(p−value)
Recombination rate (cM/Mb)
RAD51L1
67.6
67.8
68
68.2
Position on chr14 (Mb)

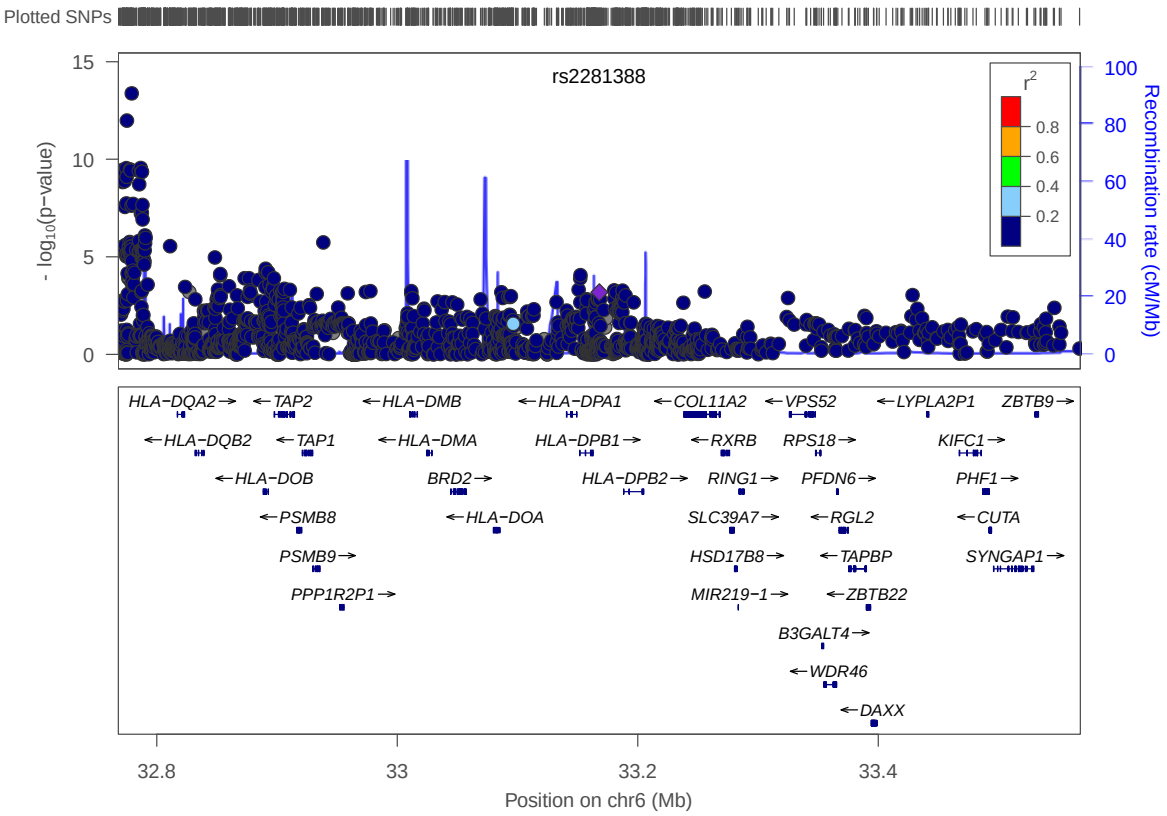

Plotted SNPs
rs4820425
r2
0.8
0.6
0.4
0.2
- log10(p-value)
Recombination rate (cM/Mb)
MKL1
SLC25A17
XPNPEP3
MIR1281
L3MBTL2
ZC3H7B
TOB2
MCHR1
ST13
RBX1
EP300
CHADL
TEF
DNAJB7
RANGAP1
Position on chr22 (Mb)

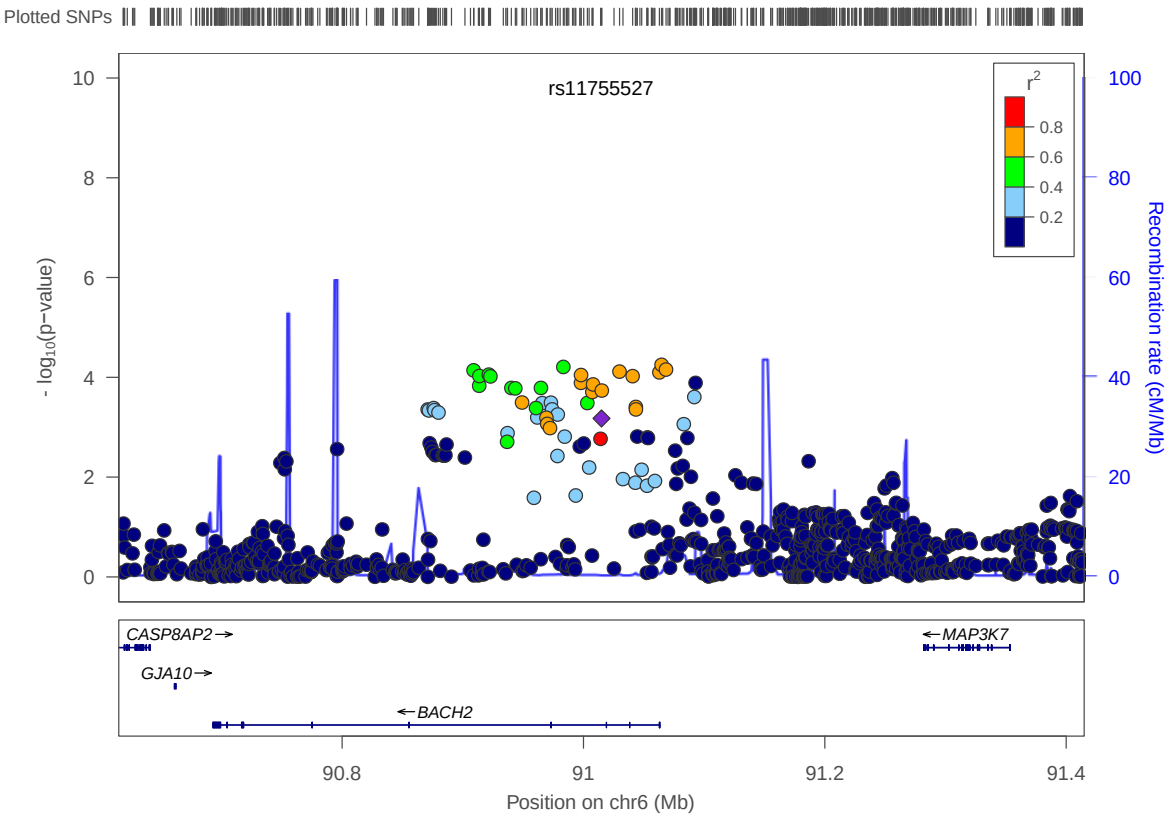

Plotted SNPs
rs4845604
r²
0.8
0.6
0.4
0.2
- log10(p−value)
Recombination rate (cM/Mb)
POGZ
CGN
SNX27
MRPL9
THEM5
S100A10
TCHHL1
HRNR
TUFT1
CELF3
LINGO4
THEM4
S100A11
RPTN
MIR554
C1orf230
RORC
TCHH
OAZ3
TDRKH
C2CD4D
LOC100132111
149.8
150
150.2
150.4
Position on chr1 (Mb)

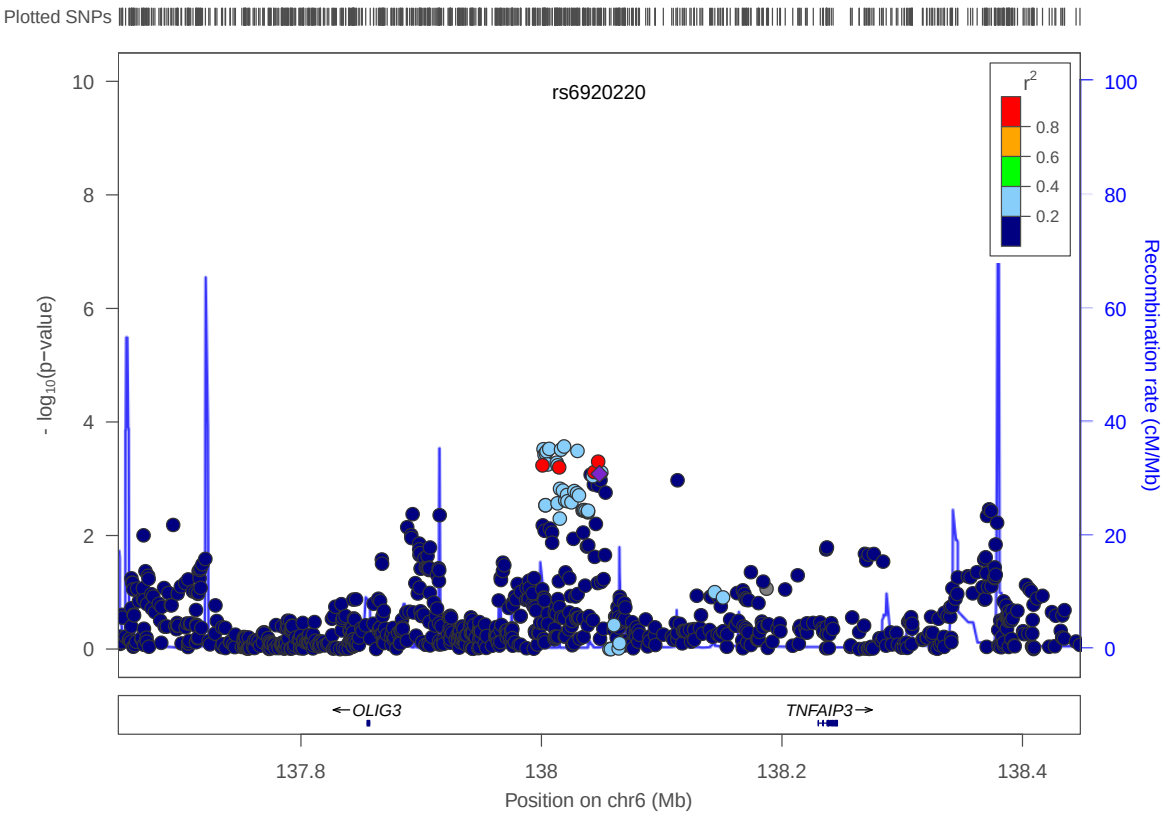

Plotted SNPs
rs12722563
r2
0.8
0.6
0.4
0.2
- log10(p-value)
Recombination rate (cM/Mb)
ASB13
GDI2
ANKRD16
IL15RA
IL2RA
PFKFB3
PRKCQ
C10orf18
FBXO18
RBM17
Position on chr10 (Mb)

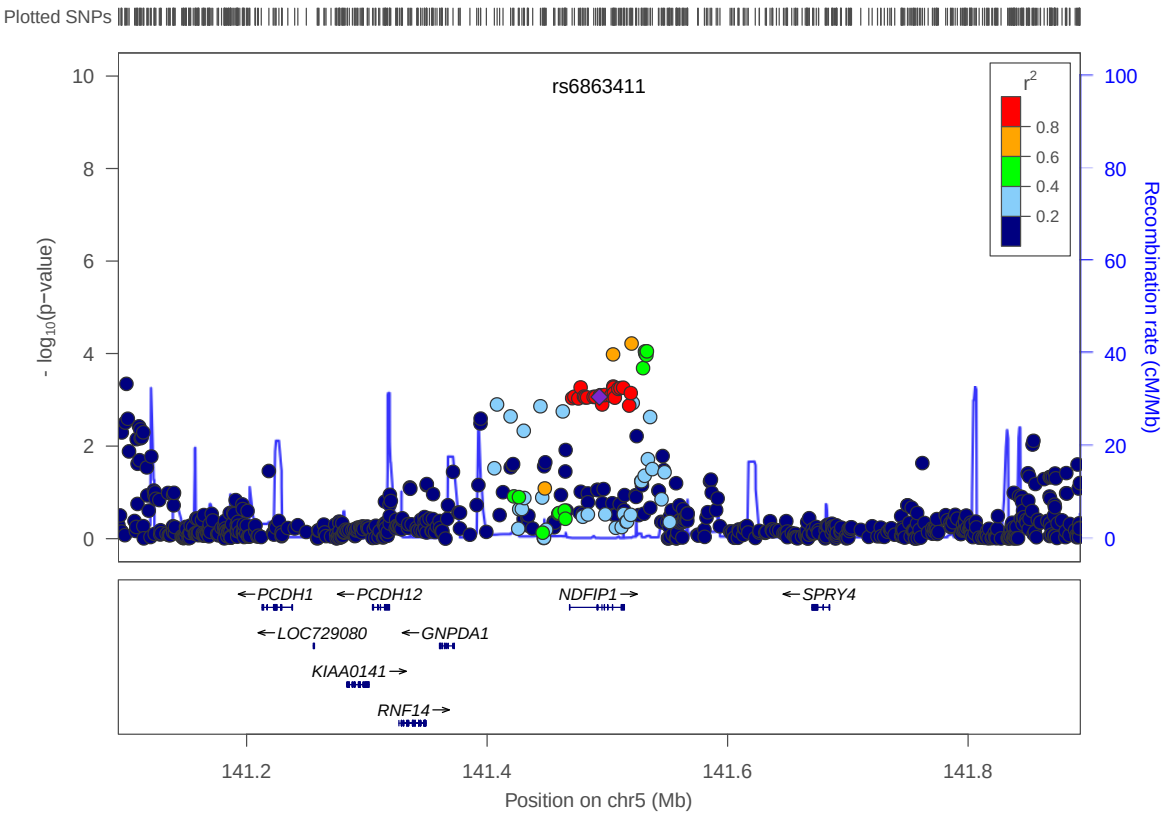

Plotted SNPs
rs2836878
r2
0.8
0.6
0.4
0.2
- log10(p−value)
Recombination rate (cM/Mb)
NCRNA00114
PSMG1
HMGN1
LCA5L
ETS2
BRWD1
WRB
SH3BGR
Position on chr21 (Mb)

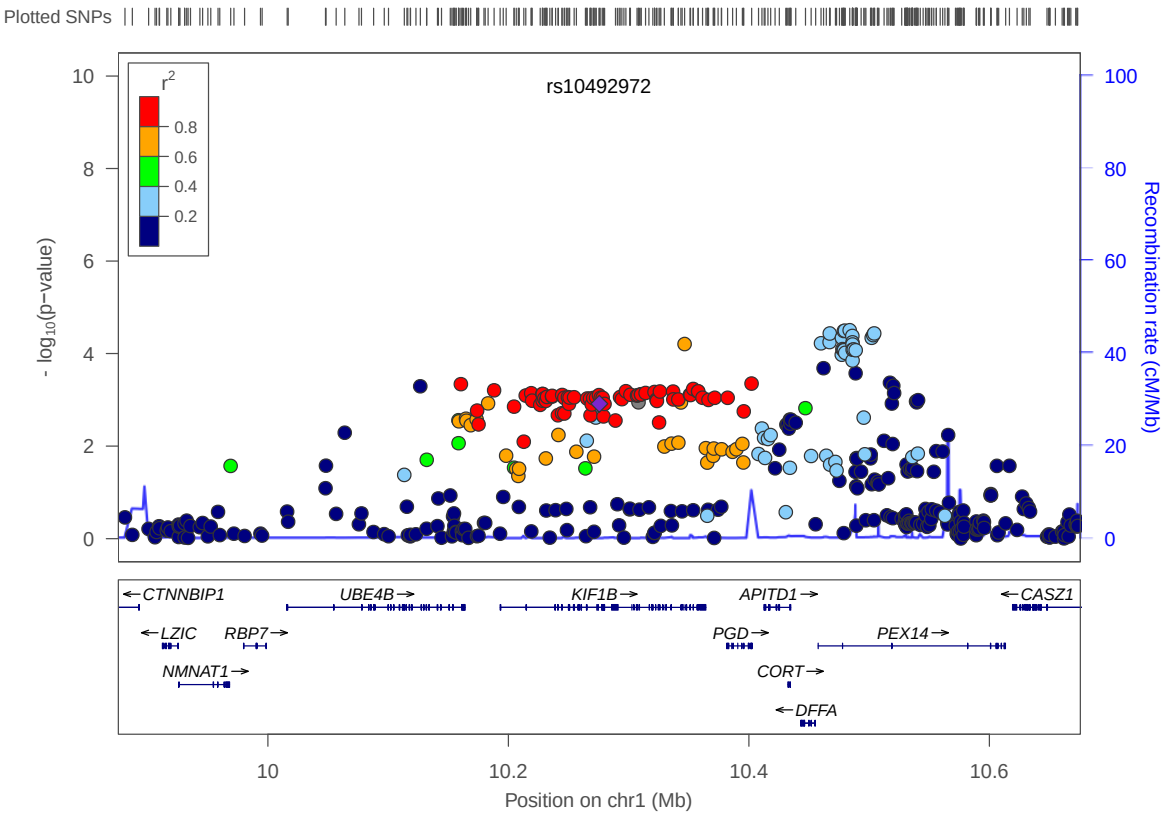

Plotted SNPs
rs11168249
r2
0.8
0.6
0.4
0.2
- log10(p-value)
Recombination rate (cM/Mb)
RPAP3
SLC48A1
VDR
TMEM106C
SENP1
ASB8
ENDOU
HDAC7
COL2A1
PFKM
RAPGEF3
C12orf68
OR10AD1
Position on chr12 (Mb)

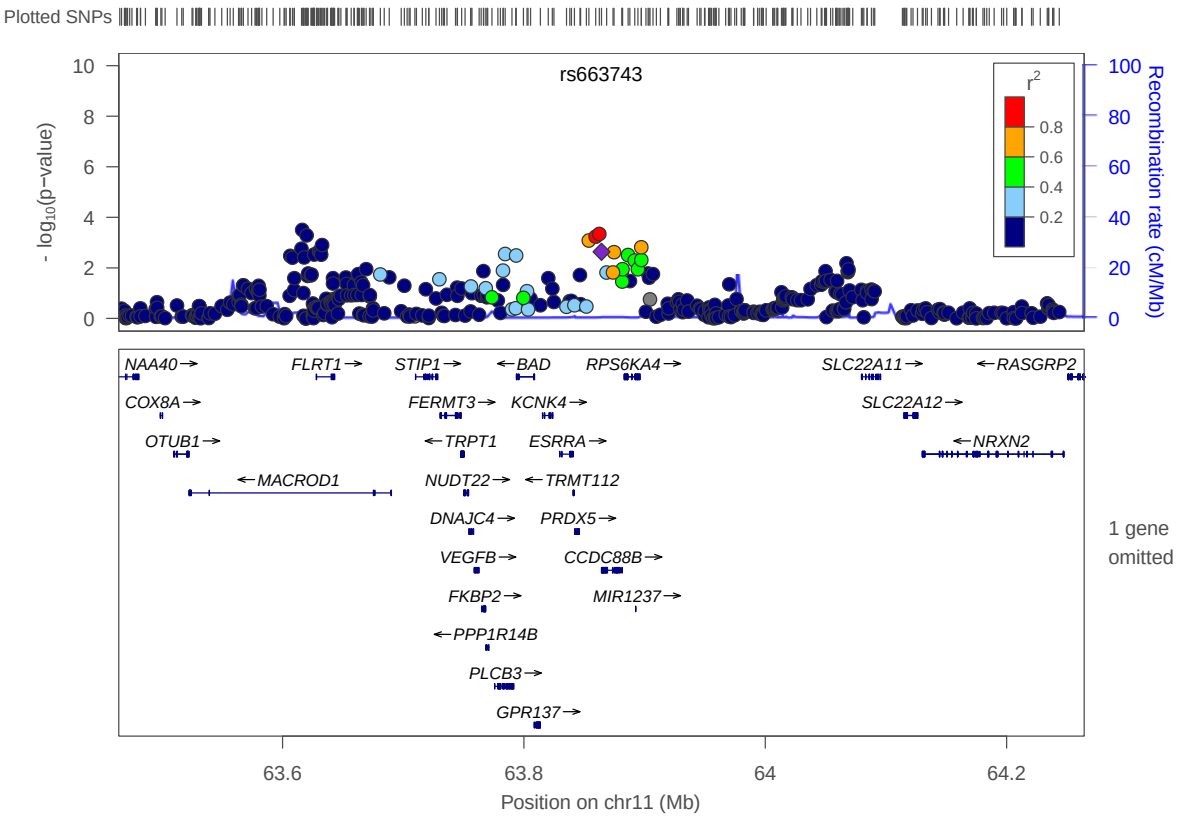

Plotted SNPs
rs2281388
r2
0.8
0.6
0.4
0.2
- log10(p-value)
Recombination rate (cM/Mb)
HLA-DQA2
TAP2
HLA-DMB
HLA-DPA1
COL11A2
VPS52
LYPLA2P1
ZBTB9
HLA-DQB2
TAP1
HLA-DMA
HLA-DPB1
RXRB
RPS18
KIFC1
HLA-DOB
BRD2
HLA-DPB2
RING1
PFDN6
PHF1
PSMB8
HLA-DOA
SLC39A7
RGL2
CUTA
PSMB9
HSD17B8
TAPBP
SYNGAP1
PPP1R2P1
MIR219-1
ZBTB22
B3GALT4
WDR46
DAXX
32.8
33
33.2
33.4
Position on chr6 (Mb)

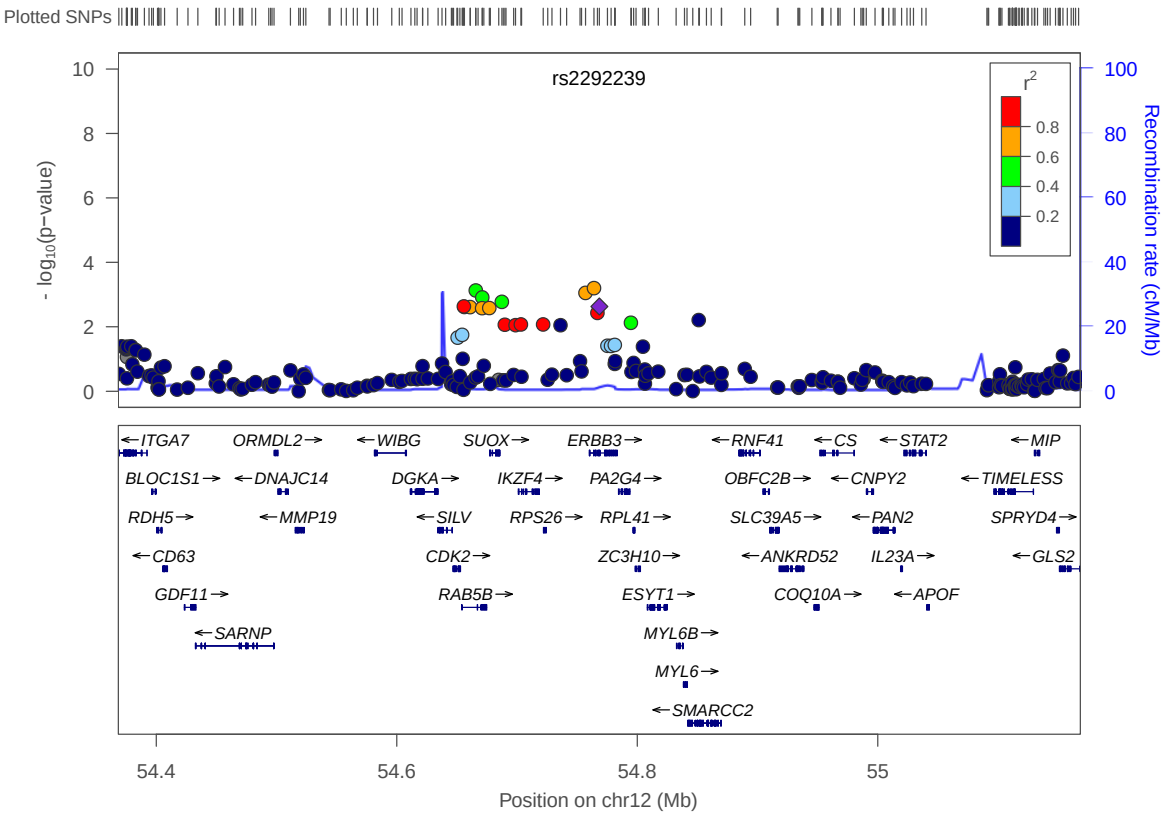

Plotted SNPs
rs11755527
r2
0.8
0.6
0.4
0.2
- log10(p−value)
Recombination rate (cM/Mb)
CASP8AP2
MAP3K7
GJA10
BACH2
Position on chr6 (Mb)

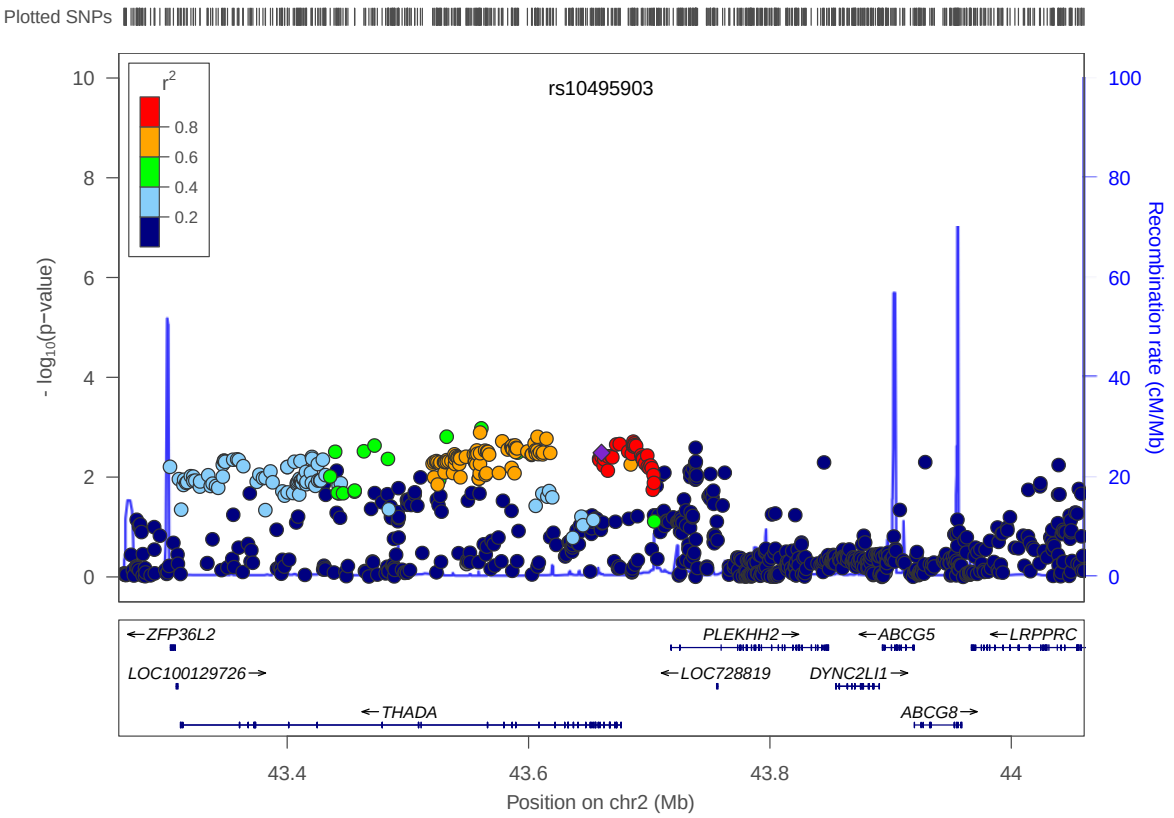

Plotted SNPs
rs6920220
r2
0.8
0.6
0.4
0.2
- log10(p−value)
Recombination rate (cM/Mb)
OLIG3
TNFAIP3
137.8
138
138.2
138.4
Position on chr6 (Mb)


Plotted SNPs
rs6863411
r2
0.8
0.6
0.4
0.2
- log10(p−value)
Recombination rate (cM/Mb)
PCDH1
PCDH12
NDFIP1
SPRY4
LOC729080
GNPDA1
KIAA0141
RNF14
Position on chr5 (Mb)


Plotted SNPs
rs10492972
r2
0.8
0.6
0.4
0.2
- log10(p−value)
Recombination rate (cM/Mb)
CTNNBIP1
UBE4B
KIF1B
APITD1
CASZ1
LZIC
RBP7
PGD
PEX14
NMNAT1
CORT
DFFA
Position on chr1 (Mb)

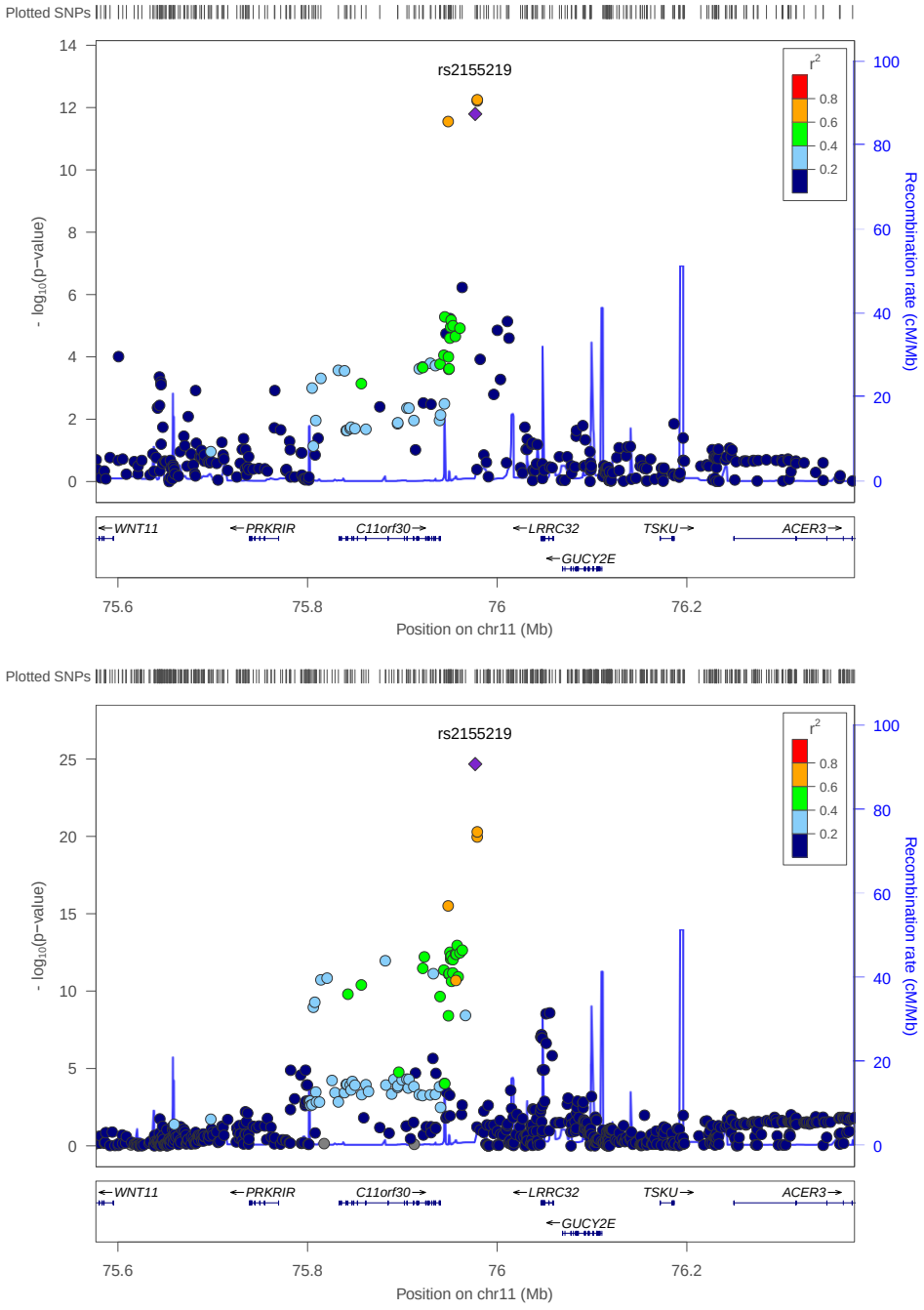

Plotted SNPs
rs663743
r2
0.8
0.6
0.4
0.2
- log10(p-value)
Recombination rate (cM/Mb)
NAA40
FLRT1
STIP1
BAD
RPS6KA4
SLC22A11
RASGRP2
COX8A
FERMT3
KCNK4
SLC22A12
OTUB1
TRPT1
ESRRA
NRXN2
MACROD1
NUDT22
TRMT112
DNAJC4
PRDX5
VEGFB
CCDC88B
FKBP2
MIR1237
PPP1R14B
PLCB3
GPR137
1 gene omitted
Position on chr11 (Mb)

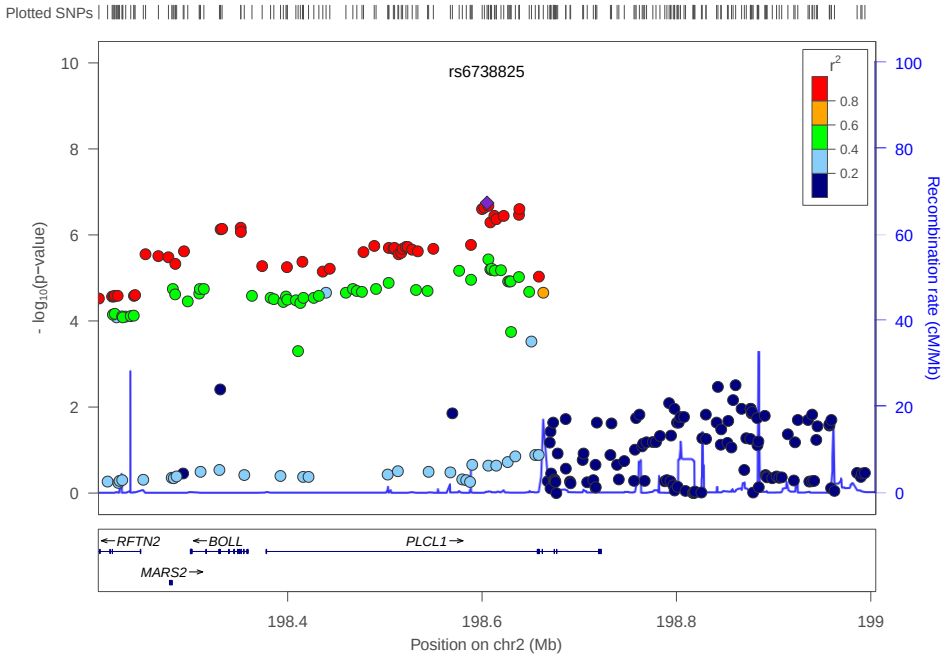

Plotted SNPs
rs2292239
r2
0.8
0.6
0.4
0.2
- log10(p−value)
Recombination rate (cM/Mb)
ITGA7
ORMDL2
WIBG
SUOX
ERBB3
RNF41
CS
STAT2
MIP
BLOC1S1
DNAJC14
DGKA
IKZF4
PA2G4
OBFC2B
CNPY2
TIMELESS
RDH5
MMP19
SILV
RPS26
RPL41
SLC39A5
PAN2
SPRYD4
CD63
CDK2
ZC3H10
ANKRD52
IL23A
GLS2
GDF11
RAB5B
ESYT1
COQ10A
APOF
SARNP
MYL6B
MYL6
SMARCC2
54.4
54.6
54.8
55
Position on chr12 (Mb)

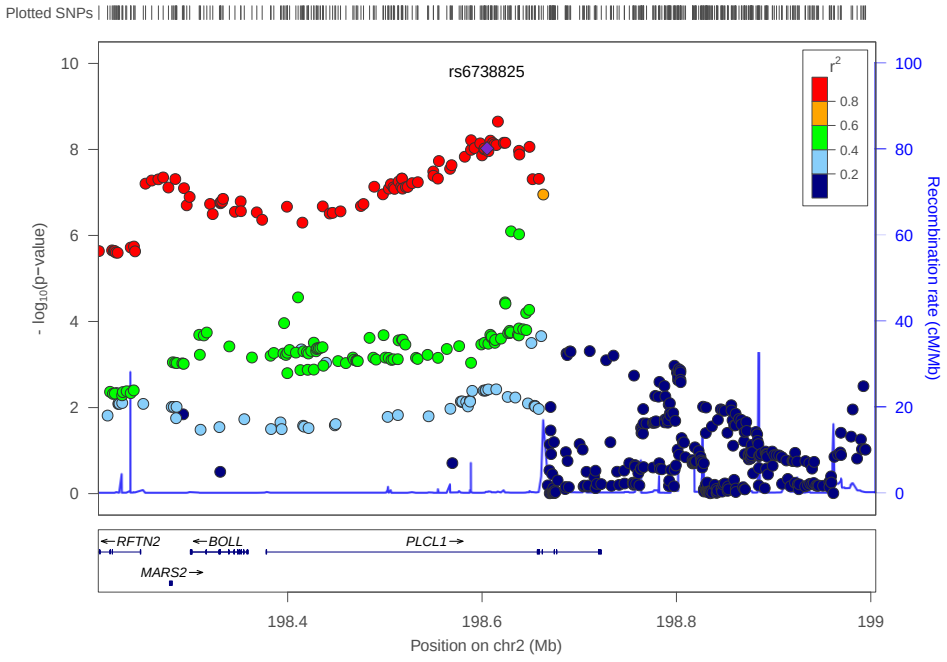

Plotted SNPs
rs10495903
r2
0.8
0.6
0.4
0.2
- log10(p−value)
Recombination rate (cM/Mb)
ZFP36L2
LOC100129726
THADA
PLEKHH2
LOC728819
DYNC2LI1
ABCG5
ABCG8
LRPPRC
43.4
43.6
43.8
44
Position on chr2 (Mb)

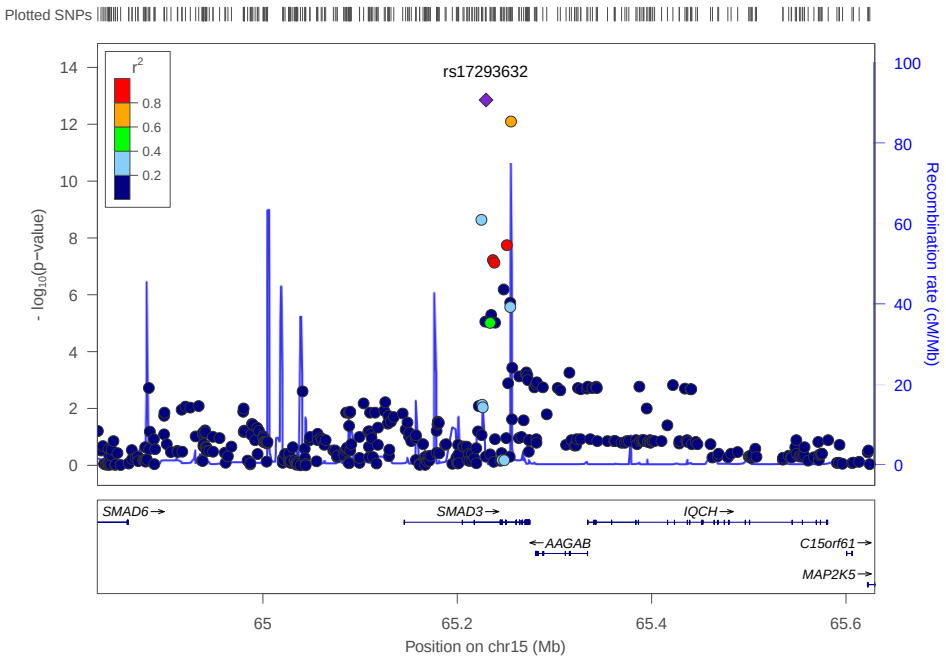
Supplementary Figure 20

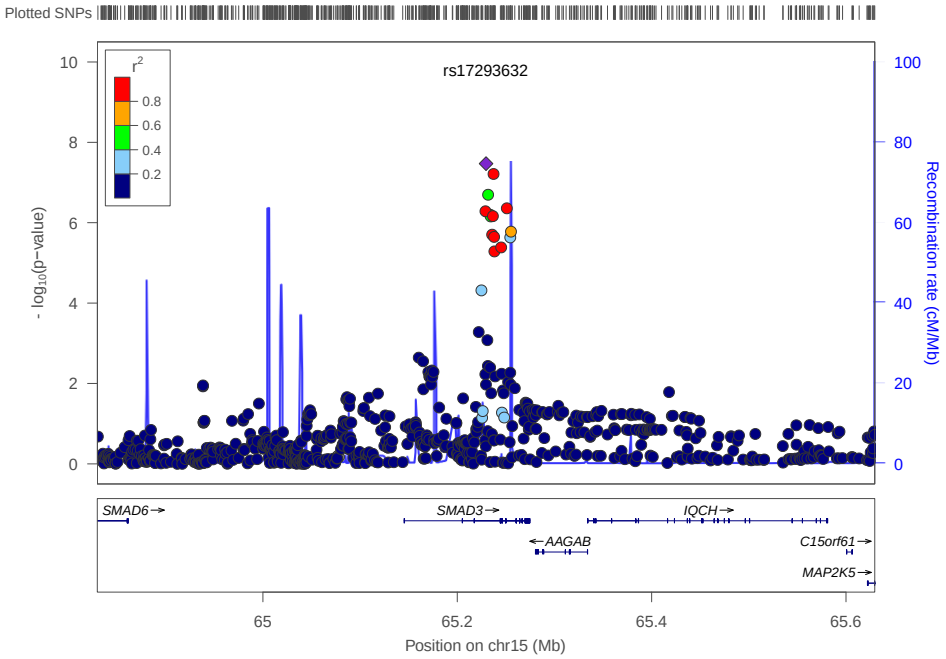
Paired LocusZoom plots within a Crohn's data[17] meta-analysis (top panel) and the allergy meta-analysis (bottom panel) for the 5 most significant shared loci.

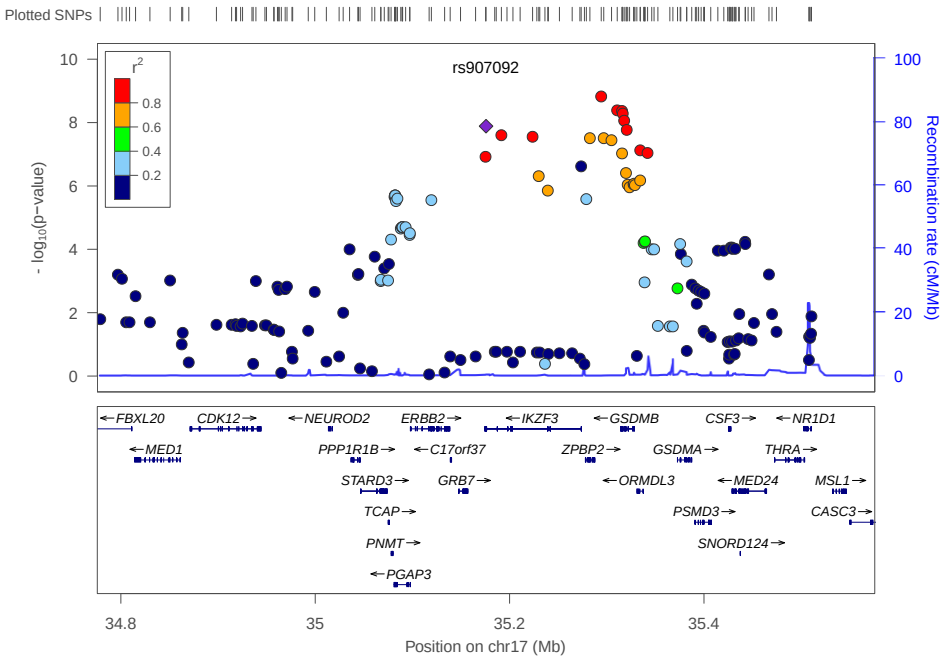


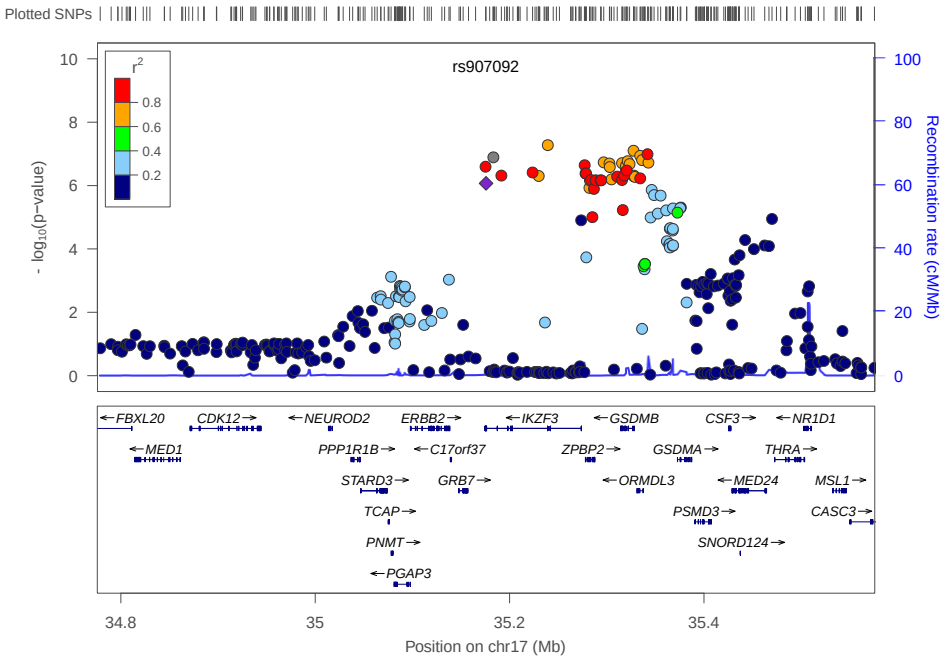

Plotted SNPs
rs6738825
r²
0.8
0.6
0.4
0.2
- log10(p-value)
Recombination rate (cM/Mb)
RFTN2
BOLL
PLCL1
MARS2
198.4
198.6
198.8
199
Position on chr2 (Mb)

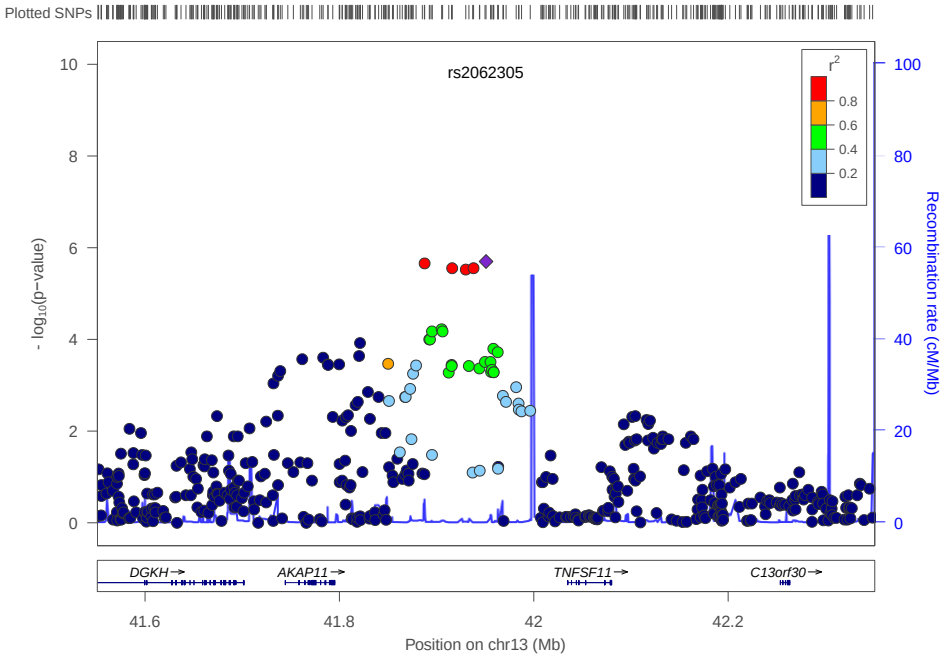

Plotted SNPs
rs6738825
r²
0.8
0.6
0.4
0.2
- log10(p-value)
Recombination rate (cM/Mb)
RFTN2
BOLL
PLCL1
MARS2
198.4
198.6
198.8
199
Position on chr2 (Mb)

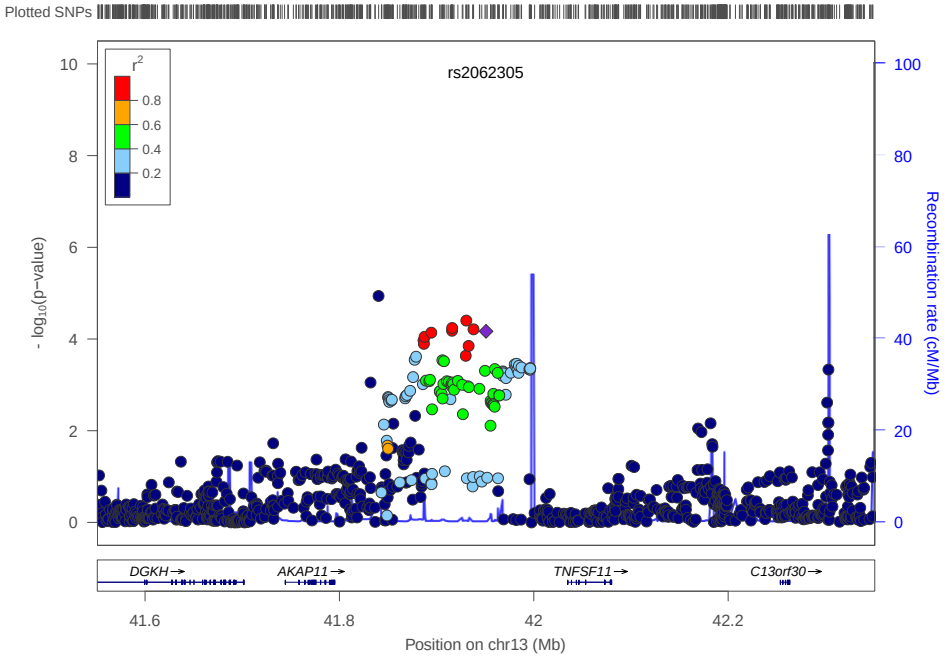

Plotted SNPs
rs17293632
r²
0.8
0.6
0.4
0.2
- log10(p−value)
Recombination rate (cM/Mb)
SMAD6
SMAD3
IQCH
AAGAB
C15orf61
MAP2K5
65
65.2
65.4
65.6
Position on chr15 (Mb)


Plotted SNPs
rs17293632
r²
0.8
0.6
0.4
0.2
- log10(p−value)
Recombination rate (cM/Mb)
SMAD6
SMAD3
IQCH
AAGAB
C15orf61
MAP2K5
65
65.2
65.4
65.6
Position on chr15 (Mb)


Plotted SNPs
rs907092
r²
0.8
0.6
0.4
0.2
- log10(p−value)
Recombination rate (cM/Mb)
FBXL20
CDK12
NEUROD2
ERBB2
IKZF3
GSDMB
CSF3
NR1D1
MED1
PPP1R1B
C17orf37
ZPBP2
GSDMA
THRA
STARD3
GRB7
ORMDL3
MED24
MSL1
TCAP
PSMD3
CASC3
PNMT
SNORD124
PGAP3
34.8
35
35.2
35.4
Position on chr17 (Mb)

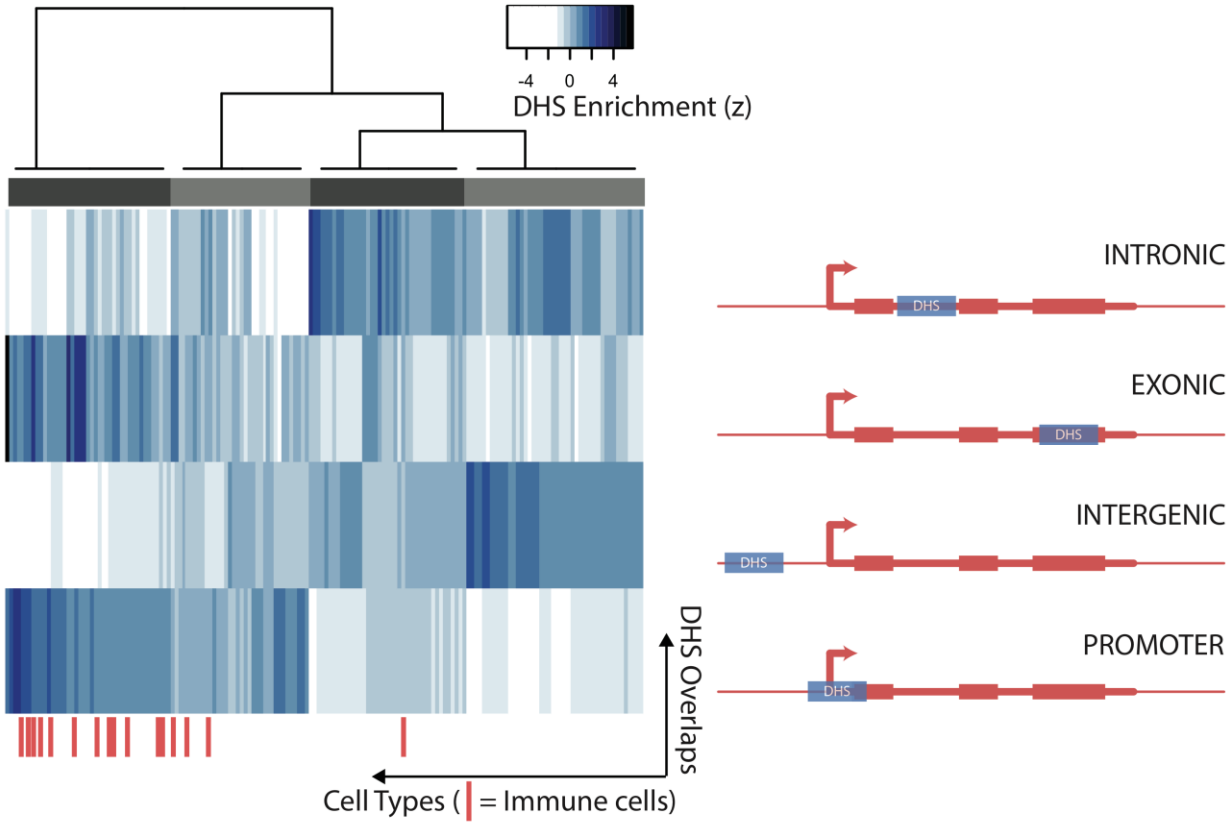

Plotted SNPs
rs907092
r²
0.8
0.6
0.4
0.2
- log10(p−value)
Recombination rate (cM/Mb)
FBXL20
CDK12
NEUROD2
ERBB2
IKZF3
GSDMB
CSF3
NR1D1
MED1
PPP1R1B
C17orf37
ZPBP2
GSDMA
THRA
STARD3
GRB7
ORMDL3
MED24
MSL1
TCAP
PSMD3
CASC3
PNMT
SNORD124
PGAP3
34.8
35
35.2
35.4
Position on chr17 (Mb)


Plotted SNPs
rs2062305
r2
0.8
0.6
0.4
0.2
- log10(p-value)
Recombination rate (cM/Mb)
DGKH
AKAP11
TNFSF11
C13orf30
41.6
41.8
42
42.2
Position on chr13 (Mb)


Plotted SNPs
rs2062305
r2
0.8
0.6
0.4
0.2
- log10(p-value)
Recombination rate (cM/Mb)
DGKH
AKAP11
TNFSF11
C13orf30
41.6
41.8
42
42.2
Position on chr13 (Mb)

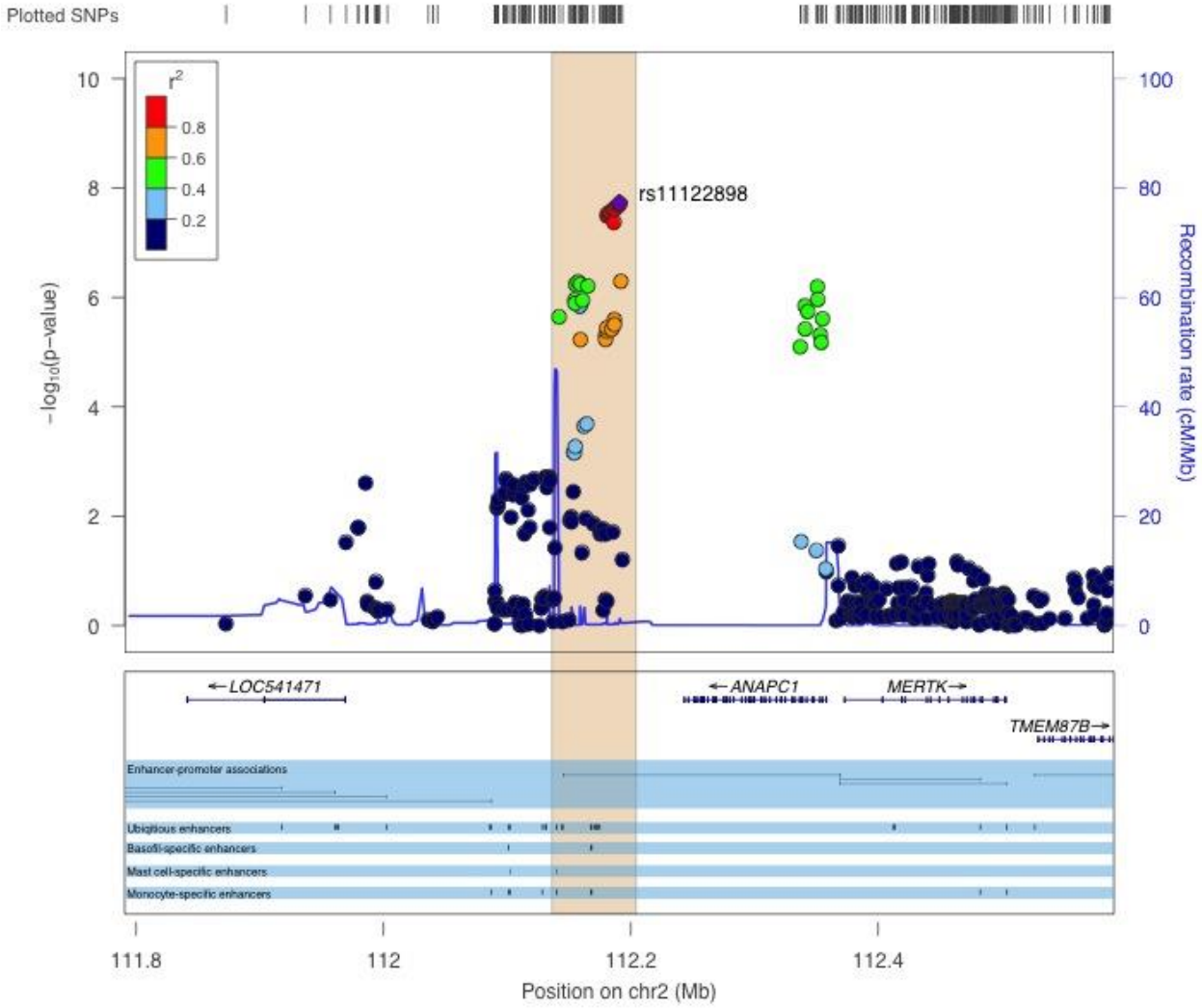
Supplementary Figure 21

ENCODE Roadmap DHS region overlap with genomic features. DHS regions for each cell type (vertical lines) were overlapped with genomic features (exons, introns, promoters, and intergenic (remaining)) (horizontal lines). Overlaps were z-scaled within each feature, and a heatmap was generated after hierarchical clustering. Immune cells are marked in red at bottom.



Supplementary Figure 22

Association plot for rs11122898 with added enhancer regions for four cell types, as well as enhancer-to gene regulatory associations, from the FANTOM5 data repository[19].



Supplement references

1. Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson J, et al. Cohort Profile: the "children of the 90s"--the index offspring of the Avon Longitudinal Study of Parents and Children. Int J Epidemiol. 2013 Feb;42(1):111–27.
2. Bønnelykke K, Matheson MC, Pers TH, Granell R, Strachan DP, Alves AC, et al. Meta-analysis of genome-wide association studies identifies ten loci influencing allergic sensitization. Nat Genet. 2013 Aug;45(8):902–6.
3. Hinds DA, McMahon G, Kiefer AK, Do CB, Eriksson N, Evans DM, et al. A genome-wide association meta-analysis of self-reported allergy identifies shared and allergy-specific susceptibility loci. Nat Genet. 2013 Aug;45(8):907–11.
4. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinforma Oxf Engl. 2010 Sep 1;26(17):2190–1.
5. Hindorff L, Junkins H, Hall P, Metha J, Manolio T. A Catalog of Published Genome-Wide Association Studies [Internet]. [cited 2011 Sep 22]. Available from: www.genome.gov/gwastudies/
6. Parkes M, Cortes A, van Heel DA, Brown MA. Genetic insights into common pathways and complex relationships among immune-mediated diseases. Nat Rev Genet [Internet]. 2013 Aug 6 [cited 2013 Aug 12];advance online publication. Available from: http://www.nature.com.ep.fjernadgang.kb.dk/nrg/journal/vaop/ncurrent/abs/nrg3502.html
7. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007 Sep;81(3):559–75.
8. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2014.
9. Consortium TEP. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012 Sep 6;489(7414):57–74.
10. Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, et al. The NIH Roadmap Epigenomics Mapping Consortium. Nat Biotechnol. 2010 Oct;28(10):1045–8.
11. Pers TH, Karjalainen JM, Chan Y, Westra HJ, Wood AR, Yang J, et al. Biological interpretation of genome-wide association studies using predicted gene functions. Nat Commun. 2015 Jan 19;6:5890.
12. Cvejic A, Haer-Wigman L, Stephens JC, Kostadima M, Smethurst PA, Frontini M, et al. SMIM1 underlies the Vel blood group and influences red blood cell traits. Nat Genet. 2013 May;45(5):542–5.
13. Lage K, Karlberg EO, Størling ZM, Olason PI, Pedersen AG, Rigina O, et al. A human phenome-interactome network of protein complexes implicated in genetic disorders. Nat Biotechnol. 2007 Mar;25(3):309–16.
14. Bult CJ, Richardson JE, Blake JA, Kadin JA, Ringwald M, Eppig JT, et al. Mouse genome informatics in a new age of biological inquiry. IEEE International Symposium on Bio-Informatics and Biomedical Engineering, 2000 Proceedings. 2000. p. 29–32.

15. Croft D, O'Kelly G, Wu G, Haw R, Gillespie M, Matthews L, et al. Reactome: a database of reactions, pathways and biological processes. Nucleic Acids Res. 2011 Jan;39(Database issue):D691–7.
16. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res. 2012 Jan;40(Database issue):D109–14.
17. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000 May;25(1):25–9.
18. Franke A, McGovern DPB, Barrett JC, Wang K, Radford-Smith GL, Ahmad T, et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. Nat Genet. 2010 Dec;42(12):1118–25.
19. Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, et al. An atlas of active enhancers across human cell types and tissues. Nature. 2014 Mar 27;507(7493):455–61.